VIRGINIA COMMONWEALTH UNIVERSITY

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations | Graduate School

2010

Interpreting the meaning of grades: A descriptive analysis of middle school teachers' assessment and grading practices

Tameshia V. Grimes
Virginia Commonwealth University

Follow this and additional works at: http://scholarscompass.vcu.edu/etd

 Part of the Education Commons

© The Author

Downloaded from
http://scholarscompass.vcu.edu/etd/2076

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

School of Education
Virginia Commonwealth University

This is to certify that the dissertation prepared by Tameshia Vaden Grimes entitled INTERPRETING THE MEANING OF GRADES: A DESCRIPTIVE ANALYSIS OF MIDDLE SCHOOL TEACHERS' ASSESSMENT AND GRADING PRACTICES has been approved by her committee as satisfactory completion of the dissertation requirement for the degree of Doctor of Philosophy

Lisa M. Abrams, Director of Dissertation

Michael D. Davis, Committee Member

Valerie J. Robnolt, Committee Member

Sherry T. Sandkam, Committee Member

Colleen A. Thoma, Director, Ph.D. in Education

Michael D. Davis, Interim Dean of the School of Education

Dr. F. Douglas Boudinot, Dean of the School of Graduate Studies

April 23, 2010

© Tameshia Vaden Grimes 2010

All Rights Reserved

INTERPRETING THE MEANING OF GRADES: A DESCRIPTIVE ANALYSIS OF MIDDLE SCHOOL TEACHERS' ASSESSMENT AND GRADING PRACTICES

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

TAMESHIA VADEN GRIMES
Ed.S., The George Washington University, 2001
M.A., Virginia Commonwealth University, 1997
B.A., Randolph-Macon College, 1996

Director: LISA M. ABRAMS
ASSISTANT PROFESSOR, SCHOOL OF EDUCATION

Virginia Commonwealth University
Richmond, Virginia
May 2010

Acknowledgements

I'd like to start by thanking Dr. Abrams, who served not only as my chairperson but also as a cheerleader and life coach during this process. She challenged my thinking and asked hard questions that forced me to re-evaluate my thoughts and face my assumptions about grading. She has encouraged me throughout the process, reminding me that I know what I am doing and am on the right path and even taunting me into thinking about possibilities beyond the dissertation. She has also been persistent in reminding me to balance the demands of my professional career, my work towards completing this doctorate, and my family time with the need to take care of myself through exercise and quiet time.

Dr. Davis, who served as my instructor for 890, deserves a special thank you. While in 890, he too asked questions that went way beyond what I envisioned for this dissertation; however, those questions helped me to find the focus and organizational structure that shaped the final dissertation. The positive experience from being in his course, as well as private conversations the two of us shared, led me to ask him to serve on my committee. He was also instrumental in suggesting others who may have been willing to 1) serve on my committee and 2) contribute to it in a meaningful way.

I also extend a warm thank you Dr. Sherry Sandkam and Dr. Valerie Robnolt for agreeing to serve on my committee without ever having met or worked with me prior to this process. The feedback each of you provided at the prospectus hearing, and the time you committed to reading my work and making comments to help make it better is greatly appreciated. The saying "it takes a village to raise a child" also applies to the writing of a

dissertation. Each of you focused on different aspects of my dissertation and gave me feedback and suggestions that were insightful and meaningful.

There have been several staff members and former co-workers along the way who have been my cheerleaders, asking about my progress and expressing their willingness to help in any way possible, especially those six teachers who came in one day during the summer to read and comment on the clarity of the survey instrument: Victoria Crone, Susan Dameron, Jenan Dahmas, Lori Hall, Janine Hoffman, and Edward Maynes.

Melinda Parker DeBrew is my best friend and most fierce advocate. She was there with an open door and an open heart when I was tired and/or discouraged. I am thankful to her for not only listening, but also for not hesitating to tell me to "get on with it and do what needs to be done."

I thank Joshua and Jasmine, my children, who tried to "have quiet time, so Mommy can do her homework." They weren't always successful, but I appreciate the fact that they tried. My desire for my children is that they remember the dedication I put into this process, and they take away from it a valuable lesson that they can apply to their approach to school: work hard and do your best, so what you learn stays with you and helps you not only reach, but also surpass, your goals.

Finally, without the support of my husband Scott, none of this would have been possible. You push me to do more than I want to do, but if it wasn't for your pushing, I wouldn't have become an administrator, never mind pursued a doctorate. Thank you. I love you, Scott, and I appreciate you more than words can express.

Table of Contents

Page

List of Tables

Page

Abstract

INTERPRETING THE MEANING OF GRADES: A DESCRIPTIVE ANALYSIS OF MIDDLE SCHOOL TEACHERS' ASSESSMENT AND GRADING PRACTICES

By Tameshia Vaden Grimes, Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2010

Major Director: Lisa M. Abrams
Assistant Professor, School of Education

This descriptive, non-experimental, quantitative study was designed to answer the broad question, "What do grades mean?" Core academic subject middle school teachers from one large, suburban school district in Virginia were administered an electronic survey that asked them to report on aspects of their grading practices and assessment methods for one class taught during the 2008-2009 school year. The survey addressed the following topics: 1) primary purposes for grades, 2) attitudes toward grading, 3) assessment method, and 4) grading practices. Additionally, the study examined the relationship between teachers' reported assessment and grading methods and student achievement.

Overall results and results disaggregated by subject area, grade level, and student ability level suggest that teachers are consistent in what they consider the primary purposes for grades. The vast majority indicated that grades should communicate student levels of mastery of content and skills. However, sizable percentages of teachers reported that they also considered non-academic indicators such as effort, attendance, and paying attention in class when determining student grades, suggesting a lack of alignment between their reported beliefs and practice. The study examined the extent to which teachers' reported grading and assessment practices were consistent with those recommended in the literature on measurement and assessment. The study findings are consistent with those of findings from previous studies suggesting that teachers engage in "hodgepodge grading," a practice which incorporates non-academic factors into student grades. The results also show that teachers use a variety of assessment methods and types of questions when measuring student achievement. The results indicate that projects, student exhibits, essays, inclusion of zeros, and extra credit were associated with higher levels of student achievement. Conversely, norm-referencing, classwork, participation, and matching were negatively correlated with student grades and test scores.

Chapter 1

Introduction

Introduction

Almost all Americans share the common experience of receiving a grade for each course taken beginning with grade school and continuing throughout their educational careers. The practice of assigning grades has been occurring in U.S. schools since the late 1800s when progress reports were first written to provide students with a summary report on their level of achievement (Hirschenbaum et al., 1971). Today, report cards are issued several times throughout the school year to provide parents and students with reports on student progress in each subject area. Students, parents, teachers, and school administrators look to these grades as a tangible way of communicating how a student is achieving in the subject area for which the grade has been assigned.

Since the late 1970s/early 1980s, state-mandated and/or high-stakes tests, which have gained favor with politicians and the public as an objective way to chart the academic progress of American students, have been widely embraced as a mechanism of accountability for students' academic achievement (Airasian, 1988). Such tests are seen by many as the ideal summative assessment because the results are reported as numbers, they are standardized – everyone within a school district gets the same number and/or same types of questions for each specific test – and the pass/fail scores are established by an

external testing authority so teachers and/or school administrators cannot change them to suit their personal professional purposes (Airasian, 1988).

Discussions about grading tend to focus on summative assessments (Airasian & Jones, 1993) – those activities, assignments, tests, etc. designed with the purpose of determining how well the student has mastered the targeted content and curriculum. However, there is an increased interest and a growing discussion about the role formative assessments can have in determining students' grades. The research on formative assessments suggests if approached and implemented correctly, formative assessments can provide the teacher with invaluable, accurate information regarding how much the student is learning while instruction is on-going (Brookhart, 2007).

Carey (1988) identified two ways school personnel use student achievement data (e.g., grades, test scores): teachers use students' performances on tests as a way to determine how they are progressing at various intervals of the school year, while school system staff members use student achievement as a measure of the effectiveness of the current curriculum and the overall performance of the school (pp. 74-75). Some parents use the grade reported to help them determine the next steps they need to take in order to help their children be successful in school (Stiggins & Knight, 1997).

Measurement experts agree that grades can be a reliable and valid tool for communicating information about a student's progress if the grade is derived from valid assessment practices and if it addresses only the area of academic achievement. For example, measurement experts recommend that information regarding non-academic behaviors, such as attendance and behavior, should be reported separately (Canady &

Hotchkiss, 1989; Friedman, 1998; McMillan, 1999; O'Connor, 1995; Wendel & Anderson, 1994). Furthermore, experts suggest that the student's grades must not be adversely impacted by grade deductions due to work that is turned in late or a perceived lack of interest in class (Canady & Hotchkiss, 1989). Despite these recommendations, researchers have found that, in practice, teachers include non-achievement factors when assigning students' grades (Baron, 2000; McMillan, 1999; McMillan & Nash, 2000). Even teachers who have received instruction in sound measurement principles engage in the practice of what Cross and Frary (1993) termed "hodgepodge grading" (Brookhart, 1993). The reasons for doing so include the desire to be "fair" to students when assessing what they know and can do (McMillan & Nash, 2000).

Statement of the Problem

Grades are universal symbols of achievement in the United States educational system; however, the processes and procedures teachers use to determine grades are not universal. Each individual teacher records and calculates grades differently; even the criteria used to determine the grade presented on the report card can vary from teacher to teacher (Carlson, 2003; Frisbie & Waltman, 1992; Guskey, 1994).

Measurement experts contend that only product criteria – summative-type assessments - should be used to determine final grades because grades issued using product criteria are based "exclusively on final examination scores, overall assessments, or other culminating demonstrations of learning" (Guskey, 1994). However, research has shown that teachers use combinations of summative and formative assessments (Brookhart, 2007), such as including/considering in the final grade the gains a student

made from the beginning of the course to the present and factoring in whether the student consistently had turned in his/her homework, along with the student's score on the final examination (Guskey, 1994, 2001). Using various types of criteria (e.g., grades given on tests, grades given for effort, or grades given for following directions) increases the chances of subjectivity and bias, invalidating the grade issued as a measure of achievement (Carlson, 2003; Frisbie & Waltman, 1992; Guskey, 1994, 2001).

Despite their agreement that grades should communicate students' academic progress, researchers disagree as to the assessment practice to be used (e.g., whether formative assessments should be used in conjunction with summative assessments or if the two should be kept strictly separate) (Black & Wiliam, 1998a, 1998b; Brookhart, 2007). Several researchers have written in support of teachers including both formative and summative information to improve classroom practices (Airasian & Jones, 1993; Black & Wiliam, 2003; Ornstein, 1994). Brookhart (2008), on the other hand, does not go as far as championing combining the two types of assessments. She does, however, state that the feedback element, which is crucial to formative assessment, can be incorporated into summative assessments, stressing the requirement that students be given the opportunity to use that information at another time to further their learning.

In the literature there is overlap between the recommendations of measurement experts regarding how best to use formative assessments to determine and communicate student achievement and how to communicate that achievement via the ultimate summative assessment: grades (Airasian & Jones, 1993; Brookhart, 2007). Experts in measurement recommend that grades be based on students' performance on criterion-

referenced activities (Carey, 1998; Cauley et al., 2008); they also recommend that formative assessments be used to provide students with feedback that has the criterion-referenced purpose of focusing students on the progress they are making towards meeting the standard/objective of the assignment (Brookhart, 2007).

Studies have shown that even when teachers are aware of the recommendations of measurement experts regarding valid and reliable grading practices, they continue to engage in grading behaviors that are contrary to those recommendations (Brookhart, 1993). The desire to be "fair" to students has been found to be the driving force behind this disregard of sound measurement principles (Airasian & Jones, 1993; McMillan & Nash, 2000). Stiggins (2002) argued that measurement experts need to move beyond a focus on validity and reliability when it comes to grading practices and begin to examine how to use grades to increase a student's desire to learn and feel successful at learning. Canady and Hotchkiss (1989) argued that teachers should be able to use professional judgment when the final, resulting grade does not accurately reflect the student's knowledge. McMillan (1999) suggested that measurement experts provide teachers with options that are less than ideal (to the specialists) but give teachers more information to better equip them to make decisions about their grading practices, resulting in grades that are a more accurate reflection of the student's learning.

Purpose of the Study

This study will examine the theory, standards, and teachers' practices of assessment and grading. The study will compare middle school, core academic subject teachers' self-reported grading and assessment practices to the recommendations of measurement experts

in an effort to determine if teachers' beliefs about the purpose of grades and if their reported assessment methods and grading practices mirror those best practices suggested by measurement experts. In addition, the relationship between grading and assessment practices and student achievement will be explored. End-of-course averages in the core academic subjects (English, mathematics, science, and social studies) as well as class averages for Virginia's state-mandated, high-stakes tests, the Standards of Learning (SOLs), will be analyzed in association with teachers' reported assessment methods and grading practices.

Rationale and Significance of the Study

Students and parents consider grades an effective and reliable way of communicating the student's academic progress in school (Brookhart, 1993). It is therefore vital that teachers clearly define their grading systems and any terms specific to that system for parents and students and consistently follow that system (Wendel & Anderson, 1994). Having a common understanding of the grading process and the terms associated with that process is necessary to ensuring that those who determine and those who receive grades share the same common understanding of what the grades mean (O'Connor, 1995). When teachers use grades for reasons other than to communicate how the student is progressing academically (e.g., using grades as a form of punishment or using grades as a source of motivation), then the grading process – and as a result, the grade – is no longer valid, especially if the teacher has not communicated to parents and students that other factors are considered when grades are assigned (Allen, 2005; Guskey, 1994, 2001; McMillan, 1999; Wendel & Anderson, 1994). McMillan (1999) stated "if

there is an understanding that the final grade reflects more than what students know and are able to do, then teachers need to be open and explicit about how much other factors, such as effort and improvement, influence the grade" (p. 8).

In this era of increased litigation and the wide acceptance of high-stakes, standardized assessments being used as the ultimate in schools' accountability for student achievement (Airasian, 1988; Stiggins, 2002), most recently articulated in the No Child Left Behind Act of 2001 (NCLB), the accuracy of grades as a symbol of students' mastery of curriculum is increasingly important. A heightened emphasis on student achievement, coupled with the reorganization of schools labeled as failing by NCLB standards, has resulted in a closer scrutiny of students' performances on standardized assessments. There are some who would argue that because grades are a more immediate and regular tool used to communicate student progress, they should be a good indicator of how students will perform on the standardized assessments. This study will analyze teachers' grading and assessment practices in relation to best practices recommended by measurement experts as valid and reliable processes for assessing students and determining grades in an effort to determine which grading practices are associated with higher levels of student achievement. It will also examine middle school, core academic subjects teachers' self-reported grading and assessment practices in relation to the recommendations asserted by grading and measurement experts as well as students' end-of-course grades and SOL scores.

Review of Literature

During the 1990s, the American Federation of Teachers (AFT), National Education Association (NEA), and the National Council on Measurement in Education (NCME) issued a joint statement and adopted professional standards that included assessment (Stiggins, 2002) (see *Appendix A*); however, assessment is still not a primary course in teacher preparation programs (Stiggins, 2002). Stiggins (2002) drew attention to the fact that few states require, as a part of the licensure process for teachers, that candidates demonstrate their ability to appropriately assess students, nor do they require school leaders to demonstrate competence in assessment. Stiggins (2002) also noted that there is not an assessment examination for licensing – along the lines of examinations to prove competence in a specific subject area – and argues that there should be. Allen (2005) stated that current teacher-preparatory programs do not focus enough on "measurement theory and [its] application to grading practices," so teachers assign grades based on how they were given grades as students in both grade school and college. Teachers incorporate factors from a combination of summative assessments and formative assessments when assigning grades (Brookhart, 2007).

Measurement experts state that only summative assessments should be used when determining grades; they further state that the use of confounding, non-academic factors, such as effort and behavior, leads to grades being subjective and biased (Guskey, 1994). Therefore, if the purpose of grades is to communicate information regarding a student's academic achievement, these practices are not in line with the fundamental principles of measurement, so the grading process – and the resulting grade – is invalid.

Parents and students view grades as a summary of the student's mastery of the content taught for each course taken (Allen, 2005). Teachers use grades not only to communicate with parents and students about the student's academic progress, but they also use grades as incentives for students to work harder, to ability-group students, and in some cases to punish students for a lack of effort or failure to follow the rules (Allen, 2005; Frisbie & Waltman, 1992; Guskey, 1994, 2000). Measurement experts agree that grades should provide both parents and students with an accurate picture of students' achievement in the content area. They disagree, however, over how to ensure that grading practices used by teachers are effective in meeting this objective; because grading is a subjective endeavor (Carlson, 2003; Frisbie & Waltman, 1992; Guskey, 1994), there is not one best way of grading that meets the needs of every student impacted by grades and every teacher assigning grades.

Grading is, by its very nature, subjective (Carlson, 2003; Frisbie & Waltman, 1992; Ornstein, 1994); individual teachers have their own, unique systems that they follow when assigning grades to students. Canady and Hotchkiss (1989) report inconsistent grading practices within a school, from teacher to teacher, and by the same teacher from one grading period to another. Marzano (2000) called for a complete change in the way educators assign grades, stating that "grades are so imprecise that they are almost meaningless" (p. 1). Grading practices such as grading on the curve and assigning zeros and other forms of using grades as punishment (Guskey, 2000) lend credence to Marzano's assertion.

Guskey (1994) found that both averaging grades and assigning zeros are inappropriate tools to use to communicate students' progress because "grading and reporting should communicate effectively what students have learned, what they can do, and whether their learning status is in line with expectations for that level" (Guskey, 1994, p. 17). Several measurement experts and researchers support Guskey's stance (Baron, 2000; Guskey, 2002; McMillan, 1999; Wendel & Anderson, 1994).

Friedman (1998) asserted "many teachers see grading and behavior management as the same thing; they are not" (p. 78). Many teachers include non-achievement factors when determining grades (Brookhart, 1994; Cross & Frary, 1999; McMillan & Nash, 2000); this practice is not in line with the sound measurement principles of validity and reliability in regards to grading: in order to be considered valid, grades should be used to communicate only one message and that message should be about academic achievement. When other factors (e.g., non-achievement factors) are included, the validity of the grade is weakened (Allen, 2005). In addition, the consideration of non-achievement factors may open the door to bias, via subjectivity and misinterpretation of students' actions, in the grading process (Carlson, 2003; Frisbie & Waltman, 1992; Guskey, 1994).

The No Child Left Behind (NCLB) legislation supports the need for teachers to be aware of assessment and grading practices and how they impact student learning. NCLB's Adequate Yearly Progress (AYP) provision has forced educators to closely examine the progress of every student via subgroup analyses, which translates to a need for teachers to focus on criterion-referenced assessments rather than norm-referenced assessments; teachers need to know how *individual* students are performing in relation to academic

goals and objectives rather than relying heavily on class averages or the norm for all classes.

Research Questions

The following research questions guided the design of the study:

1. What do middle school, core academic subject teachers report to be the purpose of grades?
2. What assessment and grading practices do middle school, core academic subject teachers use most often in their classrooms when determining students' grades?
3. Is there a relationship between assessment and grading practices used by middle school, core academic subject teachers and students' SOL scores and end of course grades?

A secondary focus of the study was to examine the extent to which middle school, core academic subject teachers' self-reported assessment and grading practices differed by subject area, grade level, and ability level of the students.

Methodology

This quantitative, descriptive study used an electronic questionnaire to survey middle school (grades 6-8), core academic subject teachers (English, mathematics, science, and social studies) in one large, suburban school district in Virginia about their beliefs regarding the purposes of grades, their attitudes towards grading, the factors they include when determining students' grades, the grading factors they use when assigning grades, and the assessment methods used in regard to one section of the classes they taught during

the 2008-2009 school year. The school district provided the researcher with the following information specific to the 2008-2009 school year: the names and teaching locations of middle school, core academic subject teachers; the end-of-course average (numeric and/or letter with a key describing the grading scale), by course section, for each middle school, core academic subject teacher; and middle school students' gender, race, grade, age, and SOL score for each test taken during the 2008-2009 school year. The researcher linked teachers to student achievement data before stripping identifying information to maintain the confidentiality of respondents.

Teachers' responses to the survey were used to answer the question "*What do grades mean?*" by focusing on how teachers' assessment and grading practices impact their process of determining grades. The researcher used middle school, core academic subject teachers' self-reported practices to determine how frequently best practices, as recommended by measurement experts, are used, which best practices are used, and whether there is a disconnect between best practice recommendations and actual processes teachers follow when assessing and grading. The results of the study were used to frame a discussion around the extent to which subject-area, grade level, and student ability level impact teachers' reported grading and assessment practices teachers use.

Limitations and Delimitations

This research study was limited to middle school core academic subject teachers in one school district. The maximum sample size was 541; however, the researcher did not know how many of the 541 teachers would not return to the district for the 2009-2010 school year and would therefore not be able to participate in the study. Depending upon

the number of survey respondents, there was the potential that the resulting response rate would be smaller than ideal.

The researcher was dependent upon grading and assessment practices that were self-reported by respondents, and the survey was designed so that respondents referenced one section of one class – the same class – as they responded to the survey questions. Although the directions were repeated at the beginning of each section of the survey, the possibility remained that some respondents may have "missed" the single section request in the directions and therefore skewed the results of the data. Also, because participants were self-reporting, the results may have been skewed due to participants misreading a question and/or providing an answer that was less than honest.

The analysis of science data using SOL scores could only be conducted for eighth grade teachers and students because the science SOL assessment is only given at the middle level in the eighth grade. Data for sixth and seventh grade science teachers was restricted to an analysis of end-of-course averages.

Summary

Grades as a communication tool to stakeholders regarding students' academic achievement have been a part of the educational system in the United States since the late 1800s. As the methods used to communicate student progress have evolved, so have the methods for determining grades; however, these methods are not in line, necessarily, with best practices as determined by measurement experts. Consequently, grading practices often do not result in an accurate communication of students' progress because other factors that impact the grade have been included.

Through this study the researcher compared real practice to best practice by answering the question "*What do grades mean?"* This study used middle school, core subject teachers' self-reported grading and assessment practices to examine how actual practice aligns with recommended best practices and the impact of those practices on student achievement.

Terminology

Although the same words are used when educators discuss grading and assessments, those words have different meanings depending upon the context in which they are used and who is using them. Below is a listing of words and phrases often associated with grading and assessment as well as the meaning of them as they relate to this study.

- Achievement - how well students master the instructional objectives (Pilcher, 1994).
- Assessment practices – the types of questions (e.g., multiple choice, true-false, etc.) and other assessment methods (e.g., portfolios, projects/reports, district assessments, state assessments, etc.) teachers have students complete in an effort to determine what students know and can do.
- Assignments – defined by Ebel and Frisbie (1991) as any "activities prescribed by the teacher primarily to allow students to demonstrate their level of competence" (p. 275); assignments may be formative or summative in nature.
- Criterion-referenced – assessments and/or grades that compare how a student performs on the learning objectives being measured by the particular assignment. Students are not compared to other students when criterion-referenced assessments are used (Cauley

et al., 2008; Ebel & Frisbie, 1991, p. 35) but are measured only on how close they come to meeting the objective (Carey, 1988, p. 348; Cauley et al., 2008).

- Feedback – information provided by the teacher to the student for the student to use to inform his/her progress towards meeting learning objectives and the next steps that need to be taken towards obtaining mastery (Brookhart, 2008).
- Formative assessment – method of evaluating students' progress *while learning is occurring*. Such assessments are not meant to be used as a final grade, but as a resource for teachers and students to determine where more instruction and/or focus is needed (Ebel & Frisbie, 1991, p. 24). Black and Wiliam (1998a) further define formative assessments as "encompassing all activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p. 2).
- Grading practices – the factors teachers include when determining students' grades and incorporate in the grade issued.
- Grading system – the symbols a teacher uses to communicate students' achievement to the student, parents, other educators, and other stakeholders. The meaning of these symbols must be clearly stated to and understood by all who receive the information (Ebel & Frisbie, 1991, p.268).
- Mastery – the student's demonstrated understanding of the content or development of the targeted skill.
- Measurement – defined by Ebel and Frisbie (1991) as "a verifiable observation of a more-less relationship" (p. 25). Measurement should be error-free, or reliable, and

outcomes should not be influenced by the feelings of the person interpreting the measurement; measurement results should be objective (others reach the same conclusion when given the same information) rather than subjective (Ebel & Frisbie, 1991, p. 25 and 116).

- Norm-referenced – assessments and/or grades that compare students to the average score and/or grades of a "normal" group of students in similar situations (Ebel & Frisbie, 1991, p. 34).
- Performance – according to Pilcher (1994), measurement experts associate performance with academic achievement: how well students master the instructional objectives.
- Reliability – grading is considered reliable when another teacher with the same information comes to a similar decision regarding the student's achievement on that test (Ebel & Frisbie, 1991, p. 76).
- Student exhibits – posters, reports, and other items created by students.
- Summative assessment – method of evaluating students' mastery of the curriculum taught. Such assessments occur at the end of the instructional unit (chapter, topic, etc.) and are intended to summarize whether the student is ready to move on to the next unit of study (Ebel & Frisbie, 1991, p. 24).
- Test – a specific activity or set of activities to be completed as a measure of the student's ability/achievement (Carey, 1988, p. 74).

- Validity – grading is considered valid when 1) it measures what is stated will be measured and 2) that measurement is accurate (Carey, 1988, p. 76; Ebel & Frisbie, 1991, p.100).

Chapter 2

Literature Review

Introduction

There have been numerous articles and books written on the topic of grading. Terwilliger (1966) posed this fundamental question about grading: "What should be the primary basis for assignment of marks?" (p. 33). He suggested that the way teachers approach grading should be changed so that the practice of grading becomes more standardized (e.g., based on specific policies and practices) (Terwilliger, 1966). Forty-three years later researchers are still trying to answer the question, "What do grades mean?", as well as reach a consensus regarding how to communicate other pertinent information – such as behavior, effort, attendance, etc. – about the students' development and/or progress. Brookhart (1994) wrote, "at present, earning and deserving grades seems to be rather narrowly defined as the fulfillment of external requirements, most of which have to do with turning in papers. Surely academic effort should be broader than that" (p. 298).

Stiggins, Frisbie, and Griswold (1989) list several pieces of information that are missing from the discussions of grading and grading practices: a deeper look at the everyday processes and routines of teachers to grade student performance, conversations about and reflections on the grading philosophies and subsequent grading process that

governs how teachers grade, and a listing of what is it that teachers actually do to determine and assign grades: "the student characteristics they measure, the measurement procedures they use, their rules of evidence, or the standards they apply" (Stiggins et al., 1989, p. 6).

This review will use the recommendations of measurement experts to organize the research discussions according to the following topics: 1) the implications of the measurement concepts of reliability and validity when it comes to both the intended and the received messages of grades, 2) the purpose of grades as a way to define what grades mean (i.e., what is the message being communicated by the grade assigned), 3) grading practices that are considered inappropriate by measurement experts, and 4) the dilemmas teachers face when assigning grades as evidenced by the disparity between recommended practice and actual practice.

Measurement Theory

Grading is not essential to learning or instruction (Frisbie & Waltman, 1992; Guskey, 1994); however, grades – symbols used to communicate a student's progress and/or mastery of a particular subject or concept – are major determinants of whether a student passes or fails. Historically, grades were the one area in which teachers had total control. Each individual teacher had his/her personal grading book with its unique symbols and codes that only the teacher understood and could translate; through grades, teachers had power. The problem with this practice is that it diminishes the meaning of the grades assigned. When the process used to determine grades is not explained to

stakeholders and/or varies from teacher to teacher, it is likely that the same symbols are used, but the message communicated by that symbol differs for each teacher.

Reliability and Validity. Allen and Lambating (2001) defined *validity* as "the accuracy of the assessment and grading procedures used by teachers" (p. 3) and defined *reliability* as "the dependability of the assessment and grading" (p. 3). Brookhart (1994) explains validity in terms of teachers' concerns about how the assigned grades are used/interpreted and the resulting consequences; she explains reliability as the consistency with which teachers stick to their stated standards for assigning grades when evaluating student's work. Friedman and Frisbie (1995) defined *validity*, citing Messick (1989), as "the extent to which certain inferences or actions are appropriate[,] based on assessment data" (p. 5). Baron (2000) defined Messick's theory of consequential validity in terms of grading as "taking into account the intended meaning of grades, the actual uses of grades, and the consequences of those uses" (p. 31). Messick (as cited in Brookhart, 1993) identified two components of validity (e.g., how the item will be used and the result of using it) and used them to create a four-way matrix that could be used to answer the question of validity (Brookhart, 1993). Wiliam and Black (1996) discuss reliability and validity in the context of classroom discussions and the inability of a teacher to know what conclusion to reach based only on student responses because of the various other factors that influence the discussion: language, context, physical conditions and/or environment, etc. They conclude that within the classroom, validity cannot be separated from reliability (Wiliam & Black, 1996).

There is value to considering the perspective of students and parents regarding the issue of including non-academic factors when assigning grades: "That grades are likely to be biased by the subjectivity teachers use in assessing and combining these ingredients may be perceived as less of a concern than discounting effort, ability, attitudes, conduct, and growth as irrelevant considerations…. [W]e have to recognize that to students, teachers, administrators, and possibly parents, there is considerable face validity to grades that include extraneous factors" (Allen, 2005, p. 70), when doing so helps the student's grade (Pilcher, 1994). When grades are "unidimensional" (Pilcher, 1994, p. 73) in nature, their meaning is clear and the message communicated is more likely to be the message received; however, when grades become a reflection of a "hodgepodge" (Cross & Frary, 1996) of factors, not only does the message communicated become distorted, but the reliability and validity associated with grades and grading also get questioned and lose their credibility.

Baron (2000) found a difference in the meaning teachers attach to the grading reference (referred to by Baron as the grading scale) as compared to what parents and students believe: teachers were operating from a criteria-referenced perspective, while parents and students were using a norm-referenced perspective (Baron, 2000). Baron (2000) describes the difference in how teachers and parents and students were interpreting grades as problematic because the difference in interpretation by senders and receivers of grades result in the validity of the grade being questionable.

Friedman and Frisbie (1995) conducted a qualitative study of report cards from elementary, middle, and high schools throughout the state of Wisconsin; a total of 216

report cards, selected through systematic sampling, were used in this study. Although they focused on the validity of report card grades, some of the resulting recommendations made by Friedman and Frisbie (1995) also apply to teacher-assigned grades, which get converted to the cumulative grade noted on the report card.

Friedman and Frisbie (1995) stated that when meanings associated with grading symbols are not consistent among teachers, then the validity of that symbol/system is called into question. Descriptions of symbols also should be provided in an effort to maximize the chances that the meaning of the symbol is correctly interpreted. Because the same symbols [i.e., A, B, C, D, F or E (excellent), S (satisfactory), or U (unsatisfactory)] have been used for a very long time in education, interpretations of such symbols may be based on past experiences and, therefore, be inaccurate or invalid (Friedman & Frisbie, 1995).

Friedman and Frisbie (1995) assert that grading scales must define for the interpreter whether grades are criterion-referenced or norm-referenced. They argue that the meaning to attribute to the grade is unknown and the validity of the grade is open for questioning when the reference for the grade is not known or specified. Guskey (2000, 2002, 2003) states that in order for grading to be fair, the teacher must explain the grading policy to the student prior to its implementation and consistently apply those described policies to all graded work (e.g., expectations for the student's works and the criteria to be used to determine the student's level of achievement).

Allen (2005) examined the validity of the practices teachers use to assign grades; he found that current practices are invalid when held against the principles of

measurement. Aside from the various factors used when assigning grades - factors based on teachers' personal preferences and belief systems - teachers also use one grade to communicate multiple messages: academic achievement, behavior, effort, participation, work ethic, etc. (Howley et al., 2001); the grade on the report card does not reflect only the student's academic achievement but includes other non-academic factors as well.

Recommendations of Measurement Experts. Schools should have a common system of grading in which each symbol's meaning is clearly defined to communicate a message about how well a student has performed regarding meeting the goals and objectives of the course (Ebel & Frisbie, 1991). The grading system, including an explanation of terms used in the grading system, should be clearly communicated to students and their parents; the grading system must also be followed consistently by the teacher (O'Connor, 1995; Wendel & Anderson, 1994).

Allen and Lambating (2001) suggest that teachers must agree on and follow through with the practice of assigning grades that accurately reflect the student's level of *academic* achievement (i.e., mastery of the content/curriculum). Non-academic factors such as effort and homework completion should not be included because they invalidate the message being communicated and result in an unreliable system of communication between the teacher and those who interpret the grade. Future teachers, current teachers, school staff development personnel, college professors, and school of education faculty need to honestly evaluate how they currently practice and/or teach assessment and grading and do so within the framework of measurement theory principles of validity and reliability. Teachers need to develop grading philosophies and create grading plans based

on the measurement principles (Frisbie & Waltman, 1992), rather than continuing to engage in practices that include non-academic achievement factors and are based on teachers' personal beliefs and values and other subjective factors; however, "simply admonishing teachers to avoid using zeros in calculating grades is not consistent with the needs of many, if not most, teachers" (McMillan, 1999).

Frisbie and Waltman (1992) propose that teachers do two things before they assign their first grades: 1) develop and understand their grading philosophies and 2) define their grading plans. Wendel and Anderson (1994) state that "the grading philosophy should stipulate the reasons for grading and marking, the standards for determining grades and marks, means for reporting grades and marks, and the uses for the grades and marks that are generated" (p. 80). Frisbie and Waltman (1992) offer nine questions teachers need to answer in order to develop their grading philosophies:

1. What meaning should each grade symbol carry?
2. What should "failure" mean?
3. What elements of performance should be incorporated in a grade?
4. How should the grades in a class be distributed?
5. What should the components be like that go into a final grade?
6. How should the components of the grade be combined?
7. What method should be used to assign grades?
8. Should borderline cases be reviewed?
9. What other factors can influence the philosophy of grading?

Once the grading philosophy has been developed, teachers should then use that philosophy to define a grading plan. Frisbie and Waltman (1992) outline a nine-step process for teachers to follow when defining their grading plans:

- Step 1: Identify and implement written district policy.
- Step 2: Decide what the meaning of each grade symbol will be.
- Step 3: Check the grade meanings against your instructional approach for logical consistency.
- Step 4: Identify evaluation variables, reporting variables, and grading variables separately.
- Step 5: Check to see what the grade distributions in your building have been like at your grade level in the subjects you teach.
- Step 6: Decide on the kinds and number of grading components needed.
- Step 7: Determine how much weight each grading component will have.
- Step 8: Determine how components will be combined to create a composite score or final grade.
- Step 9: Choose a method for assigning grades.

It is not enough for teachers to develop grading philosophies and define grading plans; they must also communicate to stakeholders (i.e., students, parents, other school personnel) how grades are determined and the information that grades convey, but they need professional development in how to do so. As a result of their study, Stiggins et al. (1989) suggest that teachers need more targeted training in the following areas:

1. "explore their own philosophical positions and the implication of those positions for the meaning of grades and their grading practices,"
2. "focus on the underlying foundation of good training – sound measurement of student achievement,"
3. "address procedures for measuring and providing feedback on non-achievement student characteristics," and
4. "address issues of the effects of grades on students at various grade levels and at different levels of past achievement" (p. 13).

According to Bailey and McTighe (1996), when teachers are using grades to communicate student progress, they should consider and define for the interpreters of the grades four factors which relate to the measurement principles of validity and reliability: 1) what information the grades are communicating, 2) what methods will be used to communicate grades, 3) why it is necessary to report grades, and 4) to whom the communication of grades is targeted. Similarly, Carey (1988) outlined six steps teachers should take when evaluating students' achievement:

1. Define and analyze the expected performance, attitude, or behavior.
2. Select evaluation criteria.
3. Develop an evaluation form and procedures.
4. Formatively evaluate the form and procedures.
5. Observe and rate students' performances or behaviors.
6. Summarize group performance and behaviors to identify areas requiring additional instruction.

Carlson (2003) asserts that the more assessments students are given and the greater the variety of the assessments given, the greater the validity of the assigned grade. All agree that teachers should have a written plan, prior to their assigning the first grade, that guides them by defining intended purpose of the grade.

Non-academic factors, such as behavior, effort, and attendance, should be reported separately from academic progress (Canady & Hotchkiss, 1989; Guskey, 2002; Terwilliger, 1966; Wendel & Anderson, 1994); otherwise, the grade becomes confounded (Friedman & Frisbie, 1995). Measurement experts agree that one grade cannot effectively communicate multiple messages (e.g., academic achievement, behavior, etc.). Pilcher (1994) uses the term "unidimensional" when referring to grades communicating only one message to stakeholders. Achievement should be the area measured and reported by the grade, and furthermore, the documented grade should make a statement about how the student is doing at the specific moment in time (Frisbie et al., 1989). If behavior and attitude are going to be measured, that information needs to be communicated in a manner that separates it from the grade for achievement (Canady & Hotchkiss, 1989; Carey, 1988).

Criterion-referenced grading is preferable to norm-referenced grading (Carey, 1988). A criterion-referenced grading system communicates how well students have mastered the objectives measured: performance is not measured based on how other students performed on the same test, which it would be considered a norm-referenced grading system (Carey, 1988; Ebel & Frisbie, 1991). Frary, Cross, and Weber (1993) conducted a study of 800 public school secondary teachers who were randomly selected from a list of all academic subject teachers in the Commonwealth of Virginia during the

spring of 1991. They received 536 responses (a response rate of 67%) to a 44-item questionnaire containing 17 items asking for factual information and 27 items asking the teachers their thoughts on grading and assessment practices. Frary et al. (1993) distinguished between "tests to measure knowledge of subject matter taught" and tests to measure "mastery of specific objectives" as the criteria for teachers to consider when answering items on the questionnaire (p. 24). As a result of the study, they suggest that rather than attempting grading practices in which scores on assessments reflect the percentage of mastery each student attained holistically, teachers should focus their efforts on ensuring that assessments result in valid and reliable rank ordering of students based on the students' individual mastery of the specific objective/skill assessed.

Guskey (1994) defines three categories of learning: *product criteria,* in which grades are determined by how students perform on summative-type assessments; *process criteria,* in which non-academic factors as well as formative-type assessments are used to determine students' grades; and *progress criteria,* which is a growth-model whereby students are assigned grades based on how much they have improved over a set amount of time. In practice, teachers typically use a combination of all three criteria; however, measurement experts recommend that only product criteria – summative-type assessments – should be used to determine student grades (Guskey, 1994). Stiggins (2002) argues that when improper assessment tools are used to determine and/or communicate information about students' academic progress, progress "may be mismeasured, day-to-day…That means that all the critically important day-to-day instructional decisions made by students, teachers, and parents may be based on misinformation about success. The result is the

misdiagnosis of student needs, students' misunderstanding of their own ability to learn, miscommunication to parents and others about student progress, and virtually no effective assessment *for learning* in classrooms" (p. 762).

Terwilliger (1966) suggested that policies to determine grades be based on the nature of the course (i.e., academic courses' grades as a representation of achievement, and performance courses' grades as a representation of some combination of achievement and ability). Terwilliger (1966) noted that attention needs to be given to the types of assignments – and the subsequent weight given to those assignments – teachers use to determine students' grades because, he argues, the type of assignment given is directly associated to the level of learning students attain. Although he does not suggest that every teacher give the same weight to the same assignment – or any other such lockstep approach to grading – Terwilliger (1966) does suggest that there be some uniformity by way of guidelines teachers use when planning assessments/evaluations, and he offers Bloom's *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain* (1956) as "the basis for a set of non-technical recommendations which would become generally available to teachers" (p. 36).

Although they are not endorsing Glaser's Basic Teaching Model (BTM), introduced in 1962, Ebel and Frisbie (1991) do refer to its components as they reinforce the need for teachers to assess what is taught and the need for grades to reflect students' mastery of the stated goals and objectives. Stiggins (1991) identified three broad categories in which to classify the way teachers assess students: 1) through students' academic achievement on both formal and informal written assignments, including tests,

quizzes, homework, classwork, etc., 2) through observations and judgments of students' work (i.e., projects) and behavior, and 3) through conversations with and/or about students as well as other subjective factors. These three broad categories include both summative and formative assessment activities.

Cross and Frary (1999) identify the following grading practices to avoid: do not consider a student's ability when determining grades; do not consider how far a student has come or how much a student has improved when assigning grades; do not include a student's negative behavior or attitude when assigning grades but convey that information using a separate process; do not allow homework to have a significant role in its impact on a student's grade; and do not consider a student's level or frequency of participation in classroom discussions when in the process of assigning students' grades.

Guskey (1994) states that grading practices used by teachers should directly relate to what they expect students to learn. Guskey (2000, 2002) further recommends that teachers give incompletes for missing work - instead of assigning low grades or giving a zero as a form of punishment - and require extra effort from the student to make up the work not completed; other measurement experts agree with this recommendation (Canady & Hotchkiss, 1989; McMillan, 1999). Guskey (2000) suggests that teachers set learning criteria (also known as learning objectives) and have grades reflect where the student stands regarding meeting the stated goals and objectives (Carlson, 20003; Guskey, 2000).

Purposes of Grading

The purpose of grades is to communicate to stakeholders accurate information regarding students' academic progress (Lambating & Allen, 2002). Although a single

grade is provided on a report card or transcript, that grade is usually derived from a series of grades from various assignments; however, to ensure the validity of the grade, only academic progress, or achievement, should be represented by the grade on the assignments, and therefore, the report card or transcript (Allen & Lambating, 2001; Canady & Hotchkiss, 1989; Carey, 1988; Howley et al., 2001; Friedman & Frisbie, 1995; Lambating & Allen, 2002; Wendel & Anderson, 1994). If other information (i.e., attitude, behavior, effort, and compliance) is to be communicated, then another avenue should be used to do so (Wendel & Anderson, 1994): each grade should communicate one message in order for that grade to be valid (Lambating & Allen, 2002).

Cross and Frary (1999) assert that teachers need to know and understand the distinctions between academic measurement, which is used to determine grades, versus informal assessments, which are used to make instructional decisions. Carey (1988) asserted "your students' grades should accurately summarize and reflect their performances during each term, semester, and year. To ensure this accuracy, you will need to consider the instructional goals and objectives covered during a term…and the proportion each test will contribute to the final grade" (pp. 5-6). Subsequent literature reflected an agreement with Carey's assertion described above.

According to Lambating and Allen (2002), progress should be determined based on how the student performed on classroom assessments; assessments should measure the student's understanding and/or mastery of the taught curriculum and/or stated objective of the lesson (Carlson, 2003). Guskey (2002) takes this a step further by arguing the only grades that should count are those earned when the student shows he has mastered the goal

and/or objective, which is then a reflection of what the student knows and can do. Terwilliger (1989) stated that a failing grade should be issued only when a student cannot demonstrate achievement at even the lowest level of the stated course criteria. Grades should not be used as a measure of how well a student meets the teacher's expectation of what it means to be "a good student" (Lambating & Allen, 2002, p. 8).

Stiggins, et al. (1989) asserted that there are several advantages to grading on achievement only, including the fact that the meaning associated with the grade is clear. Yet, it is surprising to note that in school systems where two reporting systems are used to separate information about academic progress from information on non-academic factors (e.g., letter grades are used for achievement and numbers are used for effort), the result has not been a more accurate reflection of how students are progressing (Brookhart, 1994).

Barnes (1985), in a qualitative study of student teachers and cooperating teachers, found distinct differences in how student teachers and cooperating teachers described the purpose of grades. Cooperating teachers viewed grades as a tool to communicate with parents and students; student teachers saw grades as having multiple purposes: as a motivator of students, as a communication tool, as a means to sort students into categories/tracks of classes, and as a self-evaluation tool for the teacher. The student teachers' thoughts about the uses of grades include purposes that measurement experts agree should not be considered when assigning grades. Canady and Hotchkiss (1989) are adamantly against grades being used to sort – or group – students into academic tracks and instead suggest that "sorting and selecting" be replaced with "teaching and learning" (p. 68).

When asked their opinions about the "value" of grades, academic content teachers generally felt that grades are extremely valuable (Terwiliger, 1966). Academic teachers also indicated – with consistent results – that grading is of "major importance" to a teacher's job; again, the exceptions were music and art teachers who tend to consider grading to be of "minor importance" on the continuum of teachers' responsibilities (Terwilliger, 1966). The majority of teachers responding believed that performance should carry more weight than observation when evaluating/assessing students' achievement (Terwilliger, 1966). Baron (2000) found that teachers, parents, and students believe that achievement should carry the most weight in determining students' grades.

In an effort to discern what grades mean to parents and to students, Pilcher (1994) conducted a multiple case study of six students of varying ability levels (e.g., two above average, two average, and two below average) and one parent for each student. Pilcher identified grading equations used by measurement experts, teachers, and parents to define what a grade is. Despite the different grading equations attributed to parents and teachers, Pilcher (1994) did not find a discrepancy between the way teachers grade and the meaning students attach to that grade. However, Pilcher (1994) did find differences in the messages parents and teachers take away from grades: teachers sometimes look to grades to get an idea of whether – and how often – students are completing the tasks the teacher has asked of them; parents look to grades to communicate if their children have achieved academically – met the academic standard set by the teacher. Furthermore, Pilcher (1994) found that parents did not believe a grade to be a good source of information when the grade was assigned based on factors other than achievement, yet students believed teachers

should include effort when doing so would help boost the student's grade but not when it would result in a lower grade. Baron (2000) found that students and parents viewed grades as a statement about how the student is performing in relation to other students, a norm-based reference, while teachers believed grades communicated information regarding the student's performance in relation to set standards, a criterion-based reference. Despite the difference in what the grade is interpreted to mean, both parents and teachers attempt to control student-behavior by using grades as a reward or as some type of punishment.

Practices of Grading

Grading, like teaching, is an individual process, so the components used to determine grades, and the resulting message communicated by those grades, vary from teacher to teacher. Teachers within the same building may use varied grading practices and attribute different meanings to the same letter grade (Canady & Hotchkiss, 1989). Most teachers grade based on their personal values and beliefs rather than on basic measurement principles (Allen & Lambating, 2001), and many teachers grade the way they were graded as a student and include such non-academic factors as effort and attitude, which is a practice contrary to those recommended by measurement experts (Lambating & Allen, 2002). Such subjectivity calls into question the validity, or accuracy, of the grade assigned (Carlson, 2003; Frisbie & Waltman, 1992; Guskey, 1994), and it results in a lack of consistency for parents and students. That's not to say that teachers should not take into consideration other information about students' behaviors (e.g., motivation), for such information could provide valuable information regarding the student's overall

development or progress (Lambating & Allen, 2002); however, teachers should not incorporate those non-academic factors in the final grade.

Canady and Hotchkiss (1989) state the flaw is in the process teachers have historically followed – and continue to follow today: "The standard practice, from kindergarten through college, seems to be to assign grades first and then to give feedback…When are students given the opportunity to find out what we want them to learn?" (p. 71). This question, posed by Canady and Hotchkiss (1989), speaks to the practice of teachers engaging students in formative assessments, where the focus is on providing feedback to help students improve, rather than on assigning a grade. Brookhart (2008) agreed and added that teachers actually weaken the function of formative assessment when they connect a grade with it. The purpose of formative assessment is to provide students with useful information that will help them increase their learning, while the purpose of grades is to make a statement regarding how much learning has occurred. These two functions are often at odds with each other.

Brookhart (2007) asserts that teachers are reluctant to give students assessments that will measure learning through activities that require students to use prior knowledge and apply that prior knowledge to new situations to solve problems (i.e., formative assessment at work). Instead, teachers continue to assess students using activities that do not move students beyond the stage of recalling information, usually in the form of multiple choice questions (Brookhart, 2007). These types of questions are straightforward in producing right or wrong answers, lend themselves to being graded (i.e., summative in nature), and are similar to standardized assessments that most districts are required to

administer to students to meet the accountability standards set forth in the *No Child Left Behind Act* (Brookhart, 2007).

Findings from Barnes's study (1985) indicate that cooperating teachers include information regarding students' progress and other variables ascertained during informal, formative assessments that occur in the classroom during instruction; therefore, grades reported to parents are not solely the results from an averaging of grades/scores on assignments. A possible explanation for this is teachers' adjusting grading to reduce the risk of having to experience the discomfort of explaining to a parent how a student can put forth much effort yet still fail, or explaining how a student can pass despite not putting forth any effort at all (Brookhart, 1994). Airasian and Jones (1993) present another explanation: due to the amount of time spent with students and the nature of the teacher-pupil relationship, it is difficult for teachers to be "dispassionate," assigning grades based solely on academic achievement.

Allen and Lambating (2001) conducted a quantitative study of 202 pre-service teachers (153 undergraduate students and 49 graduate students), 81 practicing high school and elementary school teachers, and 34 school of education professors. Participants were asked to read a case study before completing a survey designed to determine teachers' thoughts about grading and assessment (Allen & Lambating, 2001). The case study selected for use with the survey incorporated practices that did not adhere to the recommendations of measurement specialists, as well as those that did, within a true-to-life scenario (Allen & Lambating, 2001). Allen and Lambating (2001) found teachers' grading practices regarding the information used to determine the grade assigned to students are

often based on non-achievement factors - such as completing homework - and how well the students met the teacher's non-academic criteria (e.g., the teacher's perception of the amount of effort the student displayed). Such practices reinforce the subjective nature of grading, which goes against the measurement principles of validity and reliability, and therefore, calls into question the accuracy of the given grade (Lambating & Allen, 2002).

Barnes (1985) conducted a qualitative study of 93 student teachers and 88 cooperating teachers from two large teacher education schools, one public and one private. Data were collected using interviews, journal entries, and classroom observations. Participants were asked about their opinions on the process of grading. Answers to interview questions were analyzed to determine how much participants knew about the concepts of evaluation. Missing from the responses of both student teachers and cooperating teachers was a clear-cut method used to determine and communicate students' progress. Both groups described grading (i.e., student evaluation) as one of the most difficult aspects/responsibilities of teaching. In addition, cooperating teachers expressed concerns about the potential negative impact that low grades could have on students and the way they think, act, and feel about learning.

Pilcher (1994) found that rather than grades being representative of students' academic achievement, grades have different meanings depending on the value the receiver of the grade attaches to it. Her findings are consistent with those of other researchers (Cauley et al., 2006, 2008; Guskey, 2002, 2007; McMillan, 2001) that grades represent a "hodgepodge" of factors, including effort and ability as well as achievement. One explanation for teachers' continuing use of a "hodgepodge" grading system is that they do

so to moderate the number of failing grades in order to avoid questions from administration and/or the ire of parents (Cross & Frary, 1999). A second explanation for teachers' adding non-achievement factors, such as conduct, effort, and participation, is that this provides the teacher with a way to manage his/her classroom by using grades in an effort to control student behavior (Pilcher, 1994). Even when teachers did not explicitly grade for attitude, the processes described indicate that students' attitudes do factor into the teacher's grading system and, subsequently, the assigned grade. McMillan (2001) suggests that teachers discuss what factors are used and what the grade means in an effort to provide consistency in the meaning of grades.

A hodgepodge grading system is nothing new. Cross and Frary (1993) site a 1959 article written by sociologist Talcott Parson who found that even at the elementary levels teachers combined academic and social elements of students' achievement when assigning grades. Although they are willing to accept this practice – to a certain degree – at the elementary level, Cross and Frary (1993) postulate that secondary teachers should configure grades to be a measure of subject mastery only.

Terwilliger's (1966) study of teachers' grading practices found differences between teachers of academic and non-academic courses regarding how much consideration teachers give to certain factors when determining students' grades. Teachers of non-academic courses tend to rely more on factors such as behavior, effort, attendance, and classroom performance more so than teachers of academic courses. While teachers of academic courses give more consideration to quizzes, tests, and homework, very few indicated basing grades exclusively on students' test grades. More commonly teachers

indicated that they based grades on how well the student did in relation to the teacher's perceptions about the student's abilities (Terwilliger, 1966).

Cross and Frary (1999) conducted a study, modeled after one used in Frary et al. (1993), comparing perceptions of secondary teachers and students within one school system regarding grading practices. All middle and high school students and teachers within the school division were invited to participate; 465 teachers (226 middle school, 239 high school) responded to the teacher-specific survey, and 8664 students (4174 middle school, 4490 high school) responded to the student-specific survey. The teacher-specific survey consisted of 54 forced-choice items that asked teachers what they *thought* about grading and assessment practices and what they *did* when grading and assessing students (Cross & Frary, 1999). The student survey contained 51 forced-choice items, 14 of which were the same as those on the teacher's survey, and asked students which factors they believed teachers gave more consideration to and whether the students were happy with the way teachers graded (Cross & Frary, 1999, p. 56). Cross and Frary singled out responses of teachers of academic subjects and made their responses the focus of the study (152 middle school, 155 high school). Actual practices of teachers were often in contradiction to teachers' stated beliefs, and a "hodgepodge" system of grading – in which non-achievement factors are also considered when assigning grades – was used by a majority of teachers. Decisions teachers made about the grade to assign on the final report card were based on a variety of factors, including the student's effort, improvement, ability level, and in some cases, extenuating circumstances (e.g., a special education student who is only one or two points away from passing being given the benefit of the passing grade) (Brookhart,

1993). In addition, students who performed at lower achievement levels were least likely to support a "hodgepodge" grading system, but it is important to note that they were not totally against it (Cross & Frary, 1999).

Canady and Hotchkiss (1989) stated that teachers need to adopt grading practices that are more representative of students' learning and less punitive in nature. In practice, however, teachers implement grading practices that researchers view as barriers to the effective use of grades (Allen, 2005; Guskey, 2000). Ebel and Frisbie (1991) identified several practices in which teachers engage that lead to invalid grades: assigning grades based on neatness and grammatical correctness when this is not the objective of the assignment; assigning grades based on students' behavior, mannerisms, politeness, personalities, etc.; assigning grades based on students' participation and/or attendance; and assigning grades based on how the grade will impact students' motivation or self-esteem (p. 275). Carey (1988) refers to such practices as "confounding" practices that lead to grades not providing an accurate message regarding the student's achievement. Guskey (2000) identifies three grading practices widely used by teachers that are inappropriate: grading on the curve, using grades to motivate or punish students, and assigning zeros.

Cauley et al. (2008) examined how teachers' grading practices impact students' motivation and engagement. To gather the data, two sets of surveys were developed: one for students and one for teachers. There were a total of 4487 participants: 4278 students encompassing grades 4-12 and 209 elementary, middle, and high school teachers; the participants were from four school districts. Cauley et al. (2008) found teachers use a "hodgepodge" system of grading and specifically state that assigning zeros is not

recommended because it results in an invalid record of the student's academic ability. In addition, because different teachers use different variables and assign different values to those variables, the result is an inconsistent message communicated to parents and students about the meaning of the grade.

Brookhart (1993) conducted a study consisting of case studies/scenarios, used in a previous study by Manke and Lloyd (1990), to which participants were asked to respond regarding the grade they would assign to the student and why. Participants were certified teachers who were currently teaching and enrolled in a masters of education course at a specific university. Participants were divided into two groups: those who had taken a measurement course (40 participants) and those who had not taken a measurement course (44 participants); the median years of experience was five, and participants represented all grades (K-12). Brookhart (1993) uses Messick's *Theory of Validity*, specifically referencing his *progressive matrix*, to frame the research of this study. She places teachers' processes and justifications for assigning grades into the various quadrants of the progressive matrix. Brookhart (1993) found that a majority of teachers use grades as a payment system for work completed and/or turned in rather than as a strict reference for or indication of the student's academic performance.

Baron (2000), using Messick's theory, focused on three aspects of consequential validity as they relate to grades/grading: 1) what is included, 2) the grading scale (norm- versus criteria-referenced), and 3) the perceptions that result from grades. Teachers, parents, and students from a school district in New Jersey and high school counselors and college admissions staff who attended an annual college board meeting in 1998 were asked

to complete a questionnaire designed to gauge what grades mean to each of the groups and how much agreement there is among the groups regarding the meaning behind grades. The participants were asked to respond by focusing on a course in U.S. History when answering the questions. There were a total of 310 respondents: 60 high school teachers, 48 high school students, 41 parents of high school students, 115 high school counselors, and 46 college admissions staff members. The findings of this study indicate agreement among the groups that achievement should carry the most weight, but homework and class participation should also be included and weighted second to achievement (Baron, 2000).

Howley, Kusimo, and Parrott (2001) based a study around 52 middle school girls participating in a project sponsored by the National Science Foundation and 52 teachers from the schools these girls attended. Findings from Howley et al. (2001) showed that 1) teachers use criteria other than academic performance when assigning grades, and 2) teachers used grades to reward or punish student behavior. One concern noted in the study is the potential for grades to be a reflection of non-academic achievement factors – such as good behavior – which would result in invalid grades and students not adequately prepared for the next level in their academic careers. Measurement experts do not support grades being used in this manner and are unified in the assertion that grades should not be used to motivate, reward, or punish students. Using grades to motivate students or to punish students negatively skew the intended message of assigned grades, if grades are to communicate a student's progress.

Grading practices found to be appropriate by measurement experts include using criterion-referenced assessments and assigning grades based solely on academic

performance; grading practices not endorsed by measurement experts include the use of non-achievement factors to determine grades. Findings from Frary et al. (1993) show that teachers tend to support, agree with, and engage in grading practices that are contrary to what measurement experts suggest or support; McMillan (1999) found that teachers believe effort should be included in students' final grades. Baron (2000) concludes that there is a disconnect between teachers' beliefs about grading and the reality of their approaches to grading.

Realities of Grading

Teachers must function as both an advocate for their students and a judge of their students' progress or academic worthiness (Brookhart, 1993; Guskey, 1994). This dual role makes most teachers uncomfortable – after all, it is due to a tendency towards "altruism" - or the desire to help others - that many teachers enter the profession (Brookhart, 1993) – and offers justification for the subjectivity of grades (Brookhart, 1993). Terwilliger (1989) argues for a pass/fail system in which grades are based strictly on the results of summative assessments that "measure minimal objectives" (Brookhart, 1993).

Measurement experts recommend separating effort from achievement; however, in the reality of the classroom, this may not be possible because teachers view effort as having a direct impact on student achievement (McMillan, 1999). Achievement is a major factor when determining grades, but other factors are also considered and included in the grade provided (Brookhart, 1994). Several studies have documented that teachers' practices regarding determining grades do not match recommendations made by

measurement experts; this holds true even among teachers who have been exposed to measurement principles whether via a class or an in-service (Brookhart, 1993). Effort and achievement are often intertwined – or "confounded" – in the grade assigned (Brookhart, 1994).

Using grades for reward or punishment can result in a lack of student engagement; without engagement, learning does not occur. Brookhart (1994) states that by viewing grades as a reward/punishment system, teachers lose sight of grades as evaluations of academic achievement and instead focus on whether students have completed the tasks set forth by the teacher as requirements for the class. Pilcher (1994) asserts that when grades are used to reward or consequence, students engage in certain activities for the sole purpose of getting something or avoiding something, which could have a detrimental effect: students no longer value the process of learning and begin to do only enough to receive the reward or avoid the negative consequence. When grades are used as a form of payment for the work completed, then the meaning to be attached to that grade becomes less clear (Brookhart, 1993). According to Stiggins (1988), grades in this scenario become intertwined with classroom management, resulting in the reliability and validity of the grade becoming open to being questioned (Brookhart, 1993).

Grades as a source of motivation have often been used as justification for including non-achievement factors when assigning students' grades (Guskey, 2000); however, studies have shown that low grades cause students to lose interest in trying, and assigning zeros does not motivate students to try harder (Guskey, 2000). Assigning low grades as a form of punishment often results in the student's withdrawal from the learning process

rather than serving as a source of motivation to try harder. Teachers should not assign zeros as grades because doing so does not provide an accurate assessment of the student's knowledge: does the student truly know nothing? Guskey (2002) and McMillan (1999) have found that assigning zeros and averaging the zero with the other grades results in the final grade being negatively skewed, unless the point scale between all grades is the same (McMillan, 1999). In addition, assigning zeros may fall into the punishment category discussed above. The inclusion of zeros when determining grades could result in a less than accurate portrayal of the student's level of achievement (McMillan, 2001).

Howley, Kusimo, and Parrott (2001) state that including characteristics such as effort or ability to follow directions in a grade meant to signify academic achievement could lead to the unintended consequence of grades reflecting teachers' subconscious expectations of students based on their race and or class. McMillan (2001) noted that the use of effort may skew messages about the achievement of lower level students. Pilcher (1994) found that students' views regarding teachers' including effort when assigning grades varied depending on the situation: teachers should include effort when doing so will help boost the student's grade but not when doing so would result in a lower grade.

Guskey (2007) issued a "Student Learning Evidence Questionnaire" to participants in a summer professional development institute sponsored by states' regional educational service centers; 314 responses from public school educators in three states formed the data set for the study. Participants included superintendents, principals and assistant principals, counselors, special educators, and teachers from elementary, middle, and high school. In this study of the level of validity teachers and administrators assign to fifteen commonly

used factors considered when determining students' performance, Guskey (2007) found that both teachers and administrators ranked grades relatively low as far as their significance in determining student achievement. This finding is surprising since grades are used for several reasons other than a summary on a report card; grades are used to determine students' class rankings and grade-point averages (GPAs) as well as to determine students' acceptance into various programs and to make decisions regarding promotion or retention. Guskey (2007) concluded that perhaps grades ranked low on the list of items that speak to students' achievement because educators are aware of the confounding of achievement and effort and/or other non-academic factors in grades, which reduces the validity of the grade as a representation of academic achievement and possibly explains why there is a "discrepancy between academic students' grades and their performance on state accountability assessments" (Guskey, 2007, p. 23).

Inappropriate grading practices. Ebel and Frisbie (1991) identified practices that result in grades not being valid and/or reliable. The first issue is the lack of a clearly defined grading system: what does each individual grading symbol mean? Schools have given point or percentage equivalents to letter grades, but a clear-cut definition of what each letter grade means regarding the students' achievement is lacking (Wendel & Anderson, 1994). The second issue is the lack of a grading philosophy: most teachers grade the way they were graded and allow their personal ideals to influence their decisions regarding the grade to assign individual students (Lambating & Allen, 2002). Wendel and Anderson (1994) stated that "a grading philosophy should stipulate the reasons for grading and marking, the standards for determining grades and marks, means for reporting grades

and marks, and the uses for the grades and marks that are generated" (p. 80). Thirdly, some teachers use grades to motivate or as a method of reward or punishment (Pilcher, 1994), which clouds the message regarding achievement and may have a negative impact on positive motivation for students (Wendel & Anderson, 1994).

Stiggins, Frisbie, and Griswold (1989) conducted a study in which fifteen veteran high school teachers' grading practices were examined and compared to recommended practices retrieved from a sample of textbooks written to introduce teachers to the field of measurement. They found discrepancy between recommended practice and actual practice and highlighted the tendency of teachers to grade as they had been graded as students or in a manner similar to that of their colleagues. Using grading practices endorsed by measurement experts (i.e., basing grades solely on achievement) would result in more reliable grades as representative of student achievement (Cross & Frary, 1999); however, having had exposure to the principles of measurement did not impact teachers' actual practices of assigning grades (Brookhart, 1993; Cross & Frary, 1999).

In a study of practicing classroom teachers, Brookhart (1993) used a series of scenarios, taken from Manke and Lloyd's 1990 study, that fell into one of three categories: working to ability, missing work, and improvement. Participants were asked to select from the choices provided the action they would take, using only the information provided in the scenario, when assigning grades. Participants were then asked "Why did you make this choice?" Their written responses were analyzed based on Messick's theory of validity and using Messick's *progressive matrix* to score the responses. As a result of this study, Brookhart (1993) found that a majority of teachers use grades as a payment system for

work completed and/or turned in rather than as a strict reference for or indication of the student's academic performance. Decisions teachers make about the grade to assign on the final report card are based on a variety of factors, including the student's effort, improvement, ability level, and in some cases, extenuating circumstances (e.g., a special education student who is only one or two points away from passing given the benefit of the passing grade) (Brookhart, 1993).

More recently, McMillan and Workman (1999) conducted a two-part study of the grading practices of elementary and secondary teachers. In Part I, they analyzed the survey results of 2404 elementary, middle, and high school teachers regarding how teachers use various assessments and grading practices as well as the value they place on each. Participants were from seven school districts that are members of the Metropolitan Educational Research Consortium (MERC). Part II of the study analyzed and discussed the responses of 28 Part I participants to interview questions. The results of their study support the findings by Allen (2005) and Guskey (1994) that teachers use grades to communicate multiple pieces of information. The results of the McMillan and Workman (1999) study also support the assertions of Carlson (2003), Frisbie and Waltman (1992) and Guskey (1994) that grading is subjective: teachers consider and give weight to several different factors when assigning students grades.

A study of 1483 core subject teachers in grades 6-12 from 53 schools representing seven urban/metropolitan school districts conducted by McMillan (2001) using a questionnaire about assessment and grading practices supports findings of earlier studies: teachers use a variety of factors, including non-achievement factors, when determining

students' grades. McMillan (2001) divides these factors into four categories: "academic achievement, academic enablers, external benchmarks, and extra credit and borderline cases" (p. 78). McMillan (2001) found that teachers use grading practices inconsistent with recommendations of measurement experts because teachers view grades as more than a reporting of students' academic achievements. Teachers want grades to motivate students, but they also want to be "fair" to individual student's needs, which provides justification – in their minds – for including effort, improvement, and other non-achievement factors (McMillan, 2001).

The findings of Cauley et al. (2008) support previous studies' findings that secondary teachers state that academic achievement is the most significant factor considered when determining grades (Brookhart, 1993; Cross & Frary, 1999), but also admit to including non-achievement factors - such as effort and attentiveness in class and circumstantial information such as improvement and extra credit. Brookhart (1993) found that having taken a measurement course resulted in a difference in how teachers think about the meaning of a grade; however, the study also found that there is not a difference between teachers who had taken a course in measurement and those who had not when it came to the value placed on and the social aspects of grading. The discrepancy between the purpose of grades and the use of grades (Barnes, 1985) reinforces the call of researchers (Brookhart, 1994; Cross & Frary, 1999; Stiggins, 2002) that measurement principles be taught as a part of colleges' teacher preparatory programs.

Implications for Middle School Teachers

The Potential Role of Formative Assessment. Discussions about grading tend to be discussions about summative assessments (Airasian & Jones, 1993; Stiggins, 2002) – those activities, assignments, tests, etc. designed with the purpose of determining how well the student has mastered the targeted content and curriculum. However, there is an increased interest and a growing discussion about the role formative assessments can have in determining students' grades. The research on formative assessments suggest the following: if approached and implemented correctly, formative assessments can provide the teacher with invaluable, accurate information regarding how much the student is learning (Brookhart, 2007).

Formative assessments are those activities that students engage in as their knowledge of the content or concept is developing or *form*ing. The key to assessment being truly formative in nature is the element of feedback (Black & Wiliam, 1998a, 1998b; Brookhart, 2007;Wiliam & Black, 1996): the teacher provides feedback to the student, or the student self-assesses, so that the student is aware of the mistakes made and the steps to take to correct them and avoid making the same mistakes in the future. Without this feedback, the assessment is not formative (Brookhart, 2007).

Black and Wiliam (1998a) completed an extensive literature review on formative assessments, focusing on articles published in 1988 and after. They summarized and published findings about formative assessment, in *Inside the Black Box* (1998b), which picks up where Natriello (published a review in 1987) and Crooks (published a review in 1988) ended their reviews. Black and Wiliam (1998a) read 250 publications, labeled them using 47 different labels, and collapsed them into seven categories for the writing of the

review. The review is divided into the following categories: examples in evidence/classroom experience, assessment by teachers/current practice, students and formative assessment, strategies and tactics for teachers, systems, feedback, and prospects for the theory and practice of formative assessment. Black and Wiliam (1998a) reviewed studies that involved students in elementary schools as well as secondary schools. Some studies focused on instruction in math courses; others focused on instruction in science courses. Topics discussed included self-assessment, mastery learning, motivation theory, cognitive evaluation theory, and learning goals versus mastery goals. They concluded that formative assessments have a positive impact on the achievement of low achieving students and raises the achievement of all students overall (Black & Wiliam, 1998a, 1998b).

Brookhart (2008) describes a study by Butler and Nisan (1986) that found the most effective feedback was about the task/assignment and descriptive in nature, which resulted in an increase in the student's performance and motivation. Feedback is the critical piece to formative assessments because the purpose of feedback is to improve and/or extend learning (Brookhart, 2008). Wiliam et al. (2004) conducted a study of twenty-four teachers, evenly divided between math and science teachers, across six school districts. The teachers were provided with professional development sessions that explained formative assessments and were required to develop personal action plans for teaching using the principles of formative assessment; there were no restrictions as to what could be included in the plan, and the teachers were encouraged to "experiment" with their classes. The researchers then visited the teachers to observe them as they were teaching and to

discuss how they were implementing their action plans. Wiliam et al. (2004) found that the use of formative assessment had a positive effect on students' performance on required assessments and concluded that when using formative assessments, teachers incorporated good instructional practices which resulted in students' successes.

In a review of the literature on formative assessment, Brookhart (2007) discusses the common arguments against the use of formative assessments. She begins by pointing out that the terms *formative* and *summative* were originally used to distinguish methods of evaluating students, a process that is separate from assessing students, and she moves into a discussion of the ongoing debate as to whether formative assessments and summative assessments should be treated as two distinct, unrelated processes. Brookhart (2007) highlights the argument that formative and summative assessment activities should be separated or students will not value the learning but only have an interest in the grade. When the two are combined, the feedback – or comments - noted on the work, which is the formative aspect of the assessment, is ignored because the grade, the summative aspect, takes precedence with the student since it communicates the end result.

There is often an overlap between the use of formative and summative assessments in practice; the research has shown that the line of demarcation between formative and summative assessments is blurred in how teachers use them (Brookhart, 2007). Black and Wiliam (1998b) stated that teachers have had a hard time understanding the difference between formative and summative assessments, which has resulted in neither being used to its full potential. Both summative and formative assessments are used to communicate some message about a student's level of achievement. The difference lies in how that

information is used; formative assessments contain an element of feedback that defines the problem and offers suggestions/strategies for correcting it, while summative assessments are used to relay the end result so there is not an opportunity to fix the errors. The overlap is inevitable.

Brookhart (2007) asserted that formative and summative assessments can be used together to evaluate student learning as long as the assessments are used to gauge the individual's progress against the standards/objectives (i.e., criterion-referenced) rather than used to compare the student's progress to that of other students (i.e., norm-referenced). Stiggins (2002) takes the stance that formative and summative assessments can be used together as long as there is a balance between the two purposes: assessments *of* learning, such as mandated, high-stakes tests, and assessments *for* learning that guide teachers' instructional decisions.

According to Wiliam and Black (1996), all assessments can be summative in nature, although the reverse is not true; not all summative assessments can be used for a formative purpose. Wiliam and Black (1996) point out that teachers sometimes use former summative assignments as examples, or exemplars, of the current assignment so students have a concrete reference as to what is expected, thereby becoming formative in its use. Teachers also make formative assessment activities summative in nature [e.g., a homework assignment given to practice a developing skill that gets graded but no feedback is given (Wiliam & Black, 1996), and there is not an opportunity provided for students to learn from errors].

In their 2008 study, Cauley et al., state in the overview that teachers' grading practices have a direct relationship with whether students work towards mastery or if they work towards doing what is asked of them by the teacher. This statement is supported in the literature on formative assessment. Both Brookhart (2007) and Black and Wiliam (1998a, 1998b) include the 1998 work of Crooks in their reviews, and both cite Crooks as saying that assessments have become too focused on the subsequent grade assigned to the work, so the value to be found in the process of learning has been lost. Cauley et al. (2008) also note that criterion-referenced grading, in which student performance is measured against academic goals rather than in comparison to other students, has been found to increase student achievement. Formative assessments are integral to a criterion-referenced focus because, by its definition (Black & William, 1998; Brookhart, 2007), formative assessments provide the learners with feedback regarding their progress towards the instructional objectives. Black and Wiliam (1998b) identify three pieces of information that must be included in feedback: "recognition of the *desired goal*, evidence about *present position*, and some understanding of a *way to close the gap* between the two" (p. 143, emphasis in original)

Such assertions turn the discussion towards what measures, or assessments, are included when determining students' grades. There is an ongoing debate regarding the role of formative assessments when communicating student progress: should formative assessments be included, or should grades be only the product of a student's performance on summative assessments? (Black & Wiliam, 2003; Brookhart, 2007). Black and William (1998b) suggest that the two become aligned into one system so that they are used together

to improve classroom practices (Black and Wiliam, 2003). Although there is agreement that grades should be assigned based on several sources of data, there is a lack of agreement among teachers and measurement experts over what those data sources should be, how many to include, and how much weight they should carry (Ornstein, 1994).

Summary

Although there is agreement that multiple measures need to be included in teachers' assessments of students' achievement, there is not agreement regarding exactly what should be included as part of those measures and how much they should count (Guskey, 2007; Ornstein, 1994). Brookhart's (2007) summary of her review of the literature on formative assessments parallels the findings of researchers on grading: there is a disconnect between the recommendations of measurement experts and the actual practice of teachers. This finding is supported by Black's (2000) discussion of issues around expanding the practice of using formative assessments regularly. More research needs to be conducted in an effort to determine why there is a disconnect and how to fix it (Black, 2000; Brookhart, 2007). According to Brookhart (1994), areas that still need to be addressed via research include a) the issue of validity in reference to grades - what is included and the validity of those assessments, b) the relationship between motivation and grades, and c) the role of classroom management in the assignment of grades.

This research study will contribute to the literature by providing information specific to the grading and assessment practices of middle school, core academic subject teachers; previous studies have been conducted with secondary teachers, which include middle and high school teachers. The study will provide an analysis of the assessment

practices of middle school, core academic subject teachers in one large suburban school district and whether there is alignment between the teachers' stated beliefs about grading and their self-reported assessment methods and grading policies. This study will also include a comparison of teachers' self-reported assessment methods and grading practices to those recommended by measurement experts in an effort to determine if following the recommendations of measurement experts result in higher student achievement.

Chapter 3

Methodology

Introduction

This non-experimental, quantitative, descriptive study used an electronic questionnaire to survey middle school (grades 6-8) core academic subject teachers (English, mathematics, science, and social studies) in a large, suburban school district in Virginia about their grading and assessment practices. Survey questions were developed to obtain information about the participants' beliefs regarding the purposes of grades, their attitudes toward grading, the grading factors they considered when assigning grades, and the assessment methods they used in regard to one section of the classes they taught during the 2008-2009 school year.

The following research questions guided the research design and study procedures:

1. What do middle school, core academic subject teachers report to be the purpose of grades?
2. What assessment and grading practices do middle school, core academic subject teachers use most often in their classrooms when determining students' grades?

3. Is there a relationship between assessment and grading practices used by middle school, core academic subject teachers and students' SOL scores and end-of-course grades?

A secondary focus of the study was to examine the extent to which middle school, core academic subject teachers' self-reported assessment and grading practices differed by subject area, grade level, and student ability level.

Population

The target population for this study was middle school teachers in a large, suburban school district in Virginia. The district has a student population that is larger than 50,000 and includes over 13,000 middle school students. Less than 25% of the district's students receive free and reduced lunch; more than 20% of the district's middle school students receive free and reduced lunch. Greater than 50% of the students in the district are white, followed by black students who make up more than 25% of the population; the remainder of the students in the district are in the racial/ethnic categories of Hispanic (>5%), Asian (>2%), Unspecified (>1%), while American Indian/Alaskan Native and Native Hawaiian/Pacific Islander make up less than one percent of the student population. At the middle school level, the student demographic characteristics generally mirror those of the district overall; however, there are slightly more black students and slightly fewer students whose racial/ethnic categories are unspecified. The school division has a growing population of students who are English Speakers of Other Languages (ESOL); however, enrollment of ESOL students at the middle school level (19%) outpaces that at the district level (less than 5%).

The district employs over 4,000 teachers; more than 900 teach at the middle school level (grades 6–8), and over 600 teach at the high school level. The majority of the district's teachers are female (>80%); the same is true for the middle school staff. Greater than 80% of the district's teaching staff is white. Black and Hispanic teachers comprise about 10% of the district's teaching staff, and American Indians and Asians comprise less than 1% of the teaching staff. At the middle school level, the staff demographic profile is almost identical to that of the school district as a whole; however, there are slightly more black teachers and slightly fewer Hispanic teachers at the middle school level than across the entire school district. Of those teaching during the 2008-2009 school year, more than 13% had been teaching for three years or less, about 66% had been teaching between four and twenty-four years, and slightly more than 20% had taught for twenty-five years or more.

Sample

All middle school personnel teaching at least one section of a core academic subject during the 2008-2009 school year in the participating school district, with the exception of those teachers who are assigned to the same building as the researcher, were invited to participate in this study. There were a total of 969 middle school teachers for the 2008-2009 school year; 579 taught a core academic subject during the 2008-2009 school year. Forty-two teachers did not return to the school district for the 2009-2010 school year, and 38 of the 579 teachers were excluded from the survey because they taught in the same school with the researcher, who is an employee of the district in which the survey was conducted. Consequently, the resulting maximum sample size was 499.

Of those invited to participate, 199 responded to the survey (39.9% response rate). Table 1 provides comparisons by gender, race, and subject area frequencies for this study's sample and the school system's general population of middle school, core academic subject teachers. The results of this comparison show that the participants in this study are similar to the middle school, core academic subject teaching population for the larger school system in gender, race, and subject area taught. Therefore the results of this study can be considered representative of the broader middle school teaching population.

Table 1

Comparison of Frequencies: Sample to Population

	Sample		Population	
Demographics	*n*[a]	%	*n*	%
Gender				
Female	129	81.6	777	80.2
Male	29	18.4	192	19.8
Total	158	100.0	969	100.0
Ethnicity				
Asian	1	0.7	8	0.8
Black	10	6.3	103	10.7
Hispanic	0	0.0	11	1.1
White	143	90.5	844	87.4
Other	4	2.5	n/a[b]	n/a[b]
Total	158	100.0	966	100.0
Subject Area				
Language Arts	68	39.1	187	33.2
Mathematics	54	31.0	189	33.6
Science	25	14.4	95	16.9
Social Studies	27	15.5	92	16.3
Total	174	100.0	563	100.0

[a] Totals differ due to participants choosing not to respond to certain items.
[b] The data provided by the school system did not include a racial category of "other."

Participant Characteristics

One hundred ninety-nine of the 499 teachers invited to do so, chose to participate, and 25 started but did not complete the survey. As a result, 174 participants provided

usable responses, resulting in a usable response rate of 34.8%. Of these, not everyone responded to each question. The resulting population for this study was 81.6% female and 18.4% male. Tables 2-5 show demographic data disaggregated by subject area since participants were sampled by subject area. Table 2 lists the gender demographics in relation to the subject area taught.

Table 2

Frequency Distribution of Respondents' Gender by Subject Area

	English *n* (%)	Mathematics *n* (%)	Science *n* (%)	Social Studies *n* (%)	Total *n* (%)[a]
Female	57 (91.9)	46 (93.9)	15 (65.2)	11 (45.8)	129 (81.6)
Male	5 (8.1)	3 (6.1)	8 (34.8)	13 (54.2)	29 (18.4)
Total	62	49	23	24	158

[a] Totals differ due to participants choosing not to respond to certain items.

Respondents who described their race as *white* made up the majority of the participants (90.3%), followed by those who identified themselves as *black* (6.3%), those who labeled themselves as *other* (2.5%), and those who chose *Asian* (less than 1.0%). See Table 3 for a summary of participants' racial backgrounds.

Table 3

Frequency Distribution of Respondents' Ethnic/Racial Backgrounds by Subject Area

	English *n* (%)	Mathematics *n* (%)	Science *n* (%)	Social Studies *n* (%)	Total *n* (%)[a]
Asian	0 (0)	0 (0)	1 (4.3)	0 (0)	1 (0.7)
Black/African American	3 (4.8)	3 (6.0)	2 (8.7)	2 (8.7)	10 (6.3)
White/Caucasian	56 (90.3)	47 (94.0)	19 (82.7)	21 (91.3)	144 (90.5)
Other	3 (4.8)	0 (0)	1 (4.3)	0 (0)	4 (2.5)
Total	62	50	23	23	158

[a] Totals differ due to participants choosing not to respond to certain items.

The majority of participants taught seventh grade (36.4%), followed by eighth grde (34.1%) and sixth grade (26.5%). Table 4 provides a summary of the grade levels participants referenced when completing the survey.

Table 4

Frequency Distribution of Grade Level Representation by Subject Area

	English *n* (%)	Mathematics *n* (%)	Science *n* (%)	Social Studies *n* (%)	Total *n* (%)[a]
6th	21 (31.3)	17 (31.5)	5 (20.0)	8 (29.6)	51 (29.5)
7th	24 (35.8)	21 (38.9)	9 (36.0)	9 (33.3)	63 (36.4)
8th	22 (32.8)	16 (29.6)	11 (44.0)	10 (37.0)	59 (34.1)
Total	67	54	25	27	173

[a] Totals differ due to participants choosing not to respond to certain items.

The survey asked participants to reference one section of the course they taught during the previous school year (2008-2009) and to identify the ability level of that course as *comprehensive, honors/gifted,* or *inclusive/collaborative.* Most teachers participating in this study taught classes of students at the comprehensive ability level (49.7%); however, teachers of students in honors/gifted classes closely followed at 40.5%. Less than 10.0% of the respondents indicated teaching students in an inclusive/collaborative setting. The data on the ability levels represented in this study is presented in Table 5.

Table 5

Frequency Distribution of Ability Levels by Subject Area

	English *n* (%)	Mathematics *n* (%)	Science *n* (%)	Social Studies *n* (%)	Total *n* (%)[a]
Comprehensive	34 (50.0)	27 (50.0)	11 (44.0)	14 (53.8)	86 (49.7)
Honors/Gifted	26 (38.2)	23 (42.6)	11 (44.0)	10 (38.5)	70 (40.5)
Inclusive/ Collaborative	8 (11.8)	4 (7.4)	3 (12.0)	2 (7.7)	17 (9.8)
Total	68	54	25	26	173

[a] Totals differ due to participants choosing not to respond to certain items.

Instrumentation

Development. The survey instrument was developed from several existing instruments used by researchers in the area of teachers' classroom assessment and grading practices (Cauley et al., 2008; Frary et al., 1993; Guskey, 2002; McMillan & Workman, 1999; Wiliam et al., 2004). The instrument was reviewed to establish the clarity of each

question by a team of six middle school, core subject teachers, representing each academic core subject and each middle school grade: two English, two math, one science, and one social studies teachers with one from the sixth grade, three from the seventh grade, and two from the eighth grade. Five of the review group members were female; one was male. The survey instrument was revised and submitted to members of the Metropolitan Educational Research Consortium (MERC) of Virginia Commonwealth University (VCU) for further review.

Following guidelines according to Dillman (2007), the survey began with a question that was designed to pique participants' interest in completing the survey and that was easy for them to answer. Directions were included for each section of the survey, and each page included both a progress bar that charted the respondent's level of completion and navigation buttons to help respondents proceed through the survey. All questions were numbered; there were a total of 68 questions (see *Appendix B*).

The first sixty questions were designed to measure grading, grading practices, assessment practices, grading policies, and grading plans; the final eight questions were designed for the collection of demographic information about the class the respondent referenced when answering survey questions and about the respondents themselves. Questions were worded similarly to those used in studies conducted by Cauley et al. (2008) and Frary et al. (1993); response choices mirrored those used in previous studies on grading and assessment practices (Frary et al., 1993; Guskey, 2007).

The majority of the survey's questions were forced-choice with response choices indicated via radio button; Dillman (2007) suggests keeping the process as simple for

respondents as possible and states that drop-down boxes should be used infrequently as they result in the respondent having to take extra steps than necessary. A majority of the 68 questions on the survey instrument were designed to be answered in a likert-type format; exceptions were two questions that asked respondents to rank items in order of importance, three questions that asked respondents to respond via text (typing in their answers), three yes/no questions, and the final eight questions asking for course information and participant demographics, per Dillman's (2007) suggestions. Respondents were able to skip questions, and there was the option at the end of the survey for respondents to review their answers prior to submitting the completed survey.

Survey Administration. The survey was administered electronically, using Inquisite software, which is a software designed specifically for creating and administering survey questionnaires in a web-based environment. The survey was sent to potential participants via email with a link to the web survey included at the end of the email. Although the emails were sent via a distribution list, each participant received an email addressed specifically to him/her so as not to compromise the confidentiality of other participants. Every teacher in the participating district has a personal email account and access to a computer with which he/she can receive and respond to emails. The researcher worked with the school district's technology department staff, prior to the survey's administration, to ensure the technology filters would not block the survey link, so those teachers choosing to participate in the study would not have any problems doing so.

Surveys were emailed by the Inquisite administrator for the School of Education at Virginia Commonwealth University. A reminder email was sent to those who had not

responded within the first week. Inquisite is designed so that it can track responses via IP addresses, but only the administrator has access to that information, and it is only tracked for the purpose of sending email reminders only to those who had not yet responded.

Following the suggestion of Dillman (2007), each middle school core subject teacher received a pre-notification – via email – that briefly explained the study, asked for their participation and notified them that the actual survey would arrive in their Inbox in a few days. Two to three days later, a second email – consisting of a brief letter explaining the study and asking for their participation as well as a link to the web survey – was sent to each middle school teacher who taught a core academic subject during the 2008-2009 school year; excluded were those teachers who taught in the building with the researcher. One week after the initial email containing the link to the survey was sent, a reminder email with another link to the survey was sent to those who had not yet responded to the survey. The administration of the survey occurred November 2009.

Achievement Data Collection

The Standards of Learning (SOL) assessments are administered in reading, mathematics, and social studies for all students in grades six through eight. The science SOL is administered only to eighth grade students. Generally, all assessments are administered during a testing window, determined by the Virginia Department of Education (VDOE), in the spring of each school year; however, social studies and science classes may be taught as semester courses, with the SOL given during a testing window that has been established by VDOE specifically for the end of the semester. SOL scores are divided into three categories: *failing/below proficient* refers to scores below 400,

pass/proficient refers to scores in the 400-500 range, and *pass/advanced proficiency* refers to scores ranging from 500-600; a 600 is a perfect score. These assessments are used by local school boards and VDOE to determine how students and schools are performing compared to the expectations set forth in the No Child Left Behind (NCLB) legislation.

Staff from the participating school district's research office provided the researcher with the names and teaching locations of its middle school, core academic subject teachers during the 2008-2009 school year; the end-of-course average, represented by a letter grade; and the average of students' SOL scores for the same course section the respondent referenced when answering the survey questions. Respondents were given a unique numerical code that enabled the teacher survey data to be linked to average student achievement data while protecting the identity of the respondents. Although 174 teachers responded, the researcher was able to match student achievement to only 129 of those responses. All responses remained confidential.

Data Analysis

Missing Data. Information provided with the survey notified participants that they could skip questions they were not comfortable answering. Most participants answered all questions; however, the questions that they tended to skip were those that asked for information that would enable the researcher to link teachers' responses to student achievement data – subject and/or ability level – and those that asked for demographic information on the participants – gender, race, and/or ethnicity. In addition, the SOL average could not be linked to the end-of-course average for 23 respondents, because the way class information was entered into the school system's database did not provide

enough information to match end-of-course averages with SOL averages by specific class sections. The survey responses of these teachers were excluded for the analysis.

Overall Analysis. Data collected from the questionnaire completed by teachers, as well as the student achievement data provided by the school district, were entered into PASW Statistics 17 for data analysis. Achievement data were matched to the referent class noted in the teachers' survey responses. SOL scores and end-of-course grades for each student in each section assigned to the same teacher were collapsed into an overall average; end-of-course scores were converted to a four-point scale, with four representing the letter grade "A" and continuing down to zero representing the letter grade "F," so there was one end-of-course grade for each teacher entered into SPSS. Because teachers may teach differing course levels (honors, comprehensive, advanced placement, etc.), each survey respondent was asked to answer the survey questions based on their practices with one section of their classes; there was a place on the survey instrument for each respondent to note the level and the class period of the class for which they responded to the survey questions.

To describe teachers' reported purposes of grades, descriptive statistics identifying frequency counts, means, and standard deviations were calculated. Similar analytic procedures were used to describe the overall results of the frequency of teachers' reported grading practices. To examine the relationship between teachers' reported grading practices and student achievement, correlations were calculated. See *Appendix C* for a chart describing, by research question, the survey questions used to answer the question and the statistical test(s) run during data analysis.

Sub-group Analyses. To examine teacher results by classroom characteristics (subject area, grade level, and student ability level), frequency counts were calculated for each characteristic in the following areas:

1) the purpose of grading,
2) teachers' attitudes toward grading, and
3) teachers' assessment and grading practices.

For each calculation, the classroom characteristic functioned as the independent variable: subject area (e.g., English, mathematics, science, and social studies), grade level (e.g., 6th grade, 7th grade, and 8th grade), and ability level (e.g., comprehensive, honors/gifted, and inclusive/collaborative). Comprehensive classes contain students at varying degrees of average academic ability; honors/gifted classes are comprised of students whose academic abilities are above those of the average student; and inclusive/collaborative classes include both students of average academic ability and those who receive special education services.

Relationship to Student Achievement. Research questions one and two were analyzed using frequency data to report the number of responses per item. Research question number three asked if there was a relationship between assessment practices and grading practices. Correlations were calculated to determine if relationships existed between end-of-course averages and SOL averages and grading factors teachers considered when determining grades, as well as assessment methods used by teachers. Correlations were calculated and analyzed according to subject area, grade level, and student ability level.

Delimitations

This research study was limited to middle school, core academic subject teachers in one school district. The maximum sample size was 541; however, the researcher did not know how many of the 541 teachers would not return to the district for the 2009-2010 school year and therefore not be able to participate in the study. Due to the various studies conducted by the school division during the first quarter of the current school year, teachers were asked to participate in several surveys prior to the dissemination of this study's survey; this is a probable explanation for the resulting 39.9% response rate.

Summary

This chapter described the non-experimental, quantitative, descriptive study designed to answer the question: *What do grades mean?* Middle school, core academic subject teachers in one large, suburban school district were surveyed to determine their attitudes toward grading, the factors they included when determining students' grades, the grading policies they used when assigning grades, and the assessment practices they used in their classes. Data were collected specific to the 2008-2009 school year and also included student-specific information such as the students' end-of-course grades in core academic subjects and SOL test scores in the corresponding subject area.

Chapter 4

Results

Introduction

This study was developed and conducted with the purpose of investigating the question: W*hat do grades mean?* To examine this issue, a survey was administered to obtain middle school, core academic subject teachers' self-reported grading and assessment practices. The survey contained questions that were adapted from instruments used in previous studies on grading and/or assessment (Cauley et. al, 2008; Cross & Frary, 1999; Guskey, 2007). In an effort to determine the extent, if any, to which teachers' grading and assessment practices impact student achievement, correlations were calculated to determine if relationships exist between teachers' grading and assessment practices and students' academic achievement as represented by end-of-course grades and SOL scores.

An electronic survey was administered to all middle school (e.g., grades 6-8), core academic subject teachers (e.g., English, mathematics, science, and social studies) in a large, suburban school district in Virginia. Teachers were asked to report on the factors they included when determining students' grades as well as their use of certain types of assessments to measure students' knowledge and skills. Each section of the survey began with directions to respondents to focus on one class from the previous school year and answer all questions based on their application to that class. This chapter presents the

results of the survey and is organized by research question. Each section presents the overall survey results and then the results disaggregated by subject area, grade level, and student ability level to identify any within group differences.

Purpose of Grading

Overall Results. The first research question guiding this study asked what middle school, core academic subject teachers reported to be the purpose of grades. To determine the purpose of grades, a question on the survey asked respondents to select and rank what they believed to be the five main purposes of grading of the nine listed. Frequencies of the responses were calculated. The results are reported in Table 6, which provides the ranking results for each purpose. It also shows the total number of respondents who selected that purpose as one of their top five choices. Rank order results, which are based on percentages, are not necessarily in the same order as results listed by the total number of teachers selecting the option.

Table 6

Frequency of Respondents' Ranking of the Purpose of Grades

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Measurement of Student's Mastery of Academic Content	107 (65.6)	34 (20.9)	12 (7.4)	7 (4.3)	2 (1.2)	163
Provide Feedback to Students	49 (29.9)	89 (54.3)	19 (11.6)	4 (2.4)	3 (1.8)	164
Communicate to Parents	4 (3.0)	2 (1.5)	47 (35.1)	37 (27.6)	32 (23.9)	134
Provide Incentive to Student/Motivation	0 (0)	8 (6.8)	25 (21.2)	39 (33.1)	35 (29.7)	118
Measurement of Student's Level of Effort	3 (2.9)	18 (17.1)	27 (25.7)	31 (29.5)	18 (17.1)	105
Measurement of Student's Level of Responsibility	2 (2.9)	7 (10.3)	9 (13.2)	17 (25.0)	21 (30.9)	68
Select, Identify, or Group Students	2 (2.5)	7 (8.8)	25 (31.3)	9 (11.3)	18 (22.5)	80
Communicate to School Personnel	0 (0)	1 (2.1)	2 (4.3)	9 (19.1)	14 (29.8)	47
Evaluate School Programs	1 (2.1)	2 (4.2)	0 (0)	7 (14.6)	14 (29.2)	48

[b] Percentages calculated based on number of respondents selecting each purpose of grading.

[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Teachers agreed on two clear primary purposes of grades. Mixed opinions impacted the ranking of the remaining options. Ranking data shows that *measurement of student's mastery of academic content* was selected as the primary purpose of grading by 65.5% of teachers. The purpose chosen by 54.3% of respondents as second was *provide feedback to students*. Ranked number three by 35.1% was *communicate to parents*. Overall, motivating students ranked fourth, while measuring students' effort ranked fifth.

Results by Subject Area. In Tables 7-10, the purposes are listed in rank order based on the percentage of teachers selecting the purpose for the specific ranking. The percentages were calculated based on the total number of teachers selecting the purpose, not the total number of respondents, because each purpose functions as an independent variable; each purpose counts as a separate response item.

Teachers were consistent, across subject areas, with how they ranked the two main purposes of grades. Differences became evident when determining purposes three, four, and five; however, despite differences in ranking order, there was strong agreement among teachers in all subject areas as to the top five choices. For example, *measurement of student's mastery of academic content* was ranked number one; however, more total respondents chose *provide feedback to students* when ranking the options. Although teachers differed in how they ranked the purpose of grades, they generally agreed – based on the total number choosing each option – on the five purposes of grades.

Table 7 shows English teachers' ranking of the purposes of grades. Similar to overall results, English teachers were consistent in their choices for the top two purposes of grades. The majority selected *measurement of student's mastery of academic content* as

the primary purpose of grades (65.6%) and *provide feedback to students* (51.5%) as the second purpose. Consistent with the overall results, English teachers selected among the same options for the top five purposes of grades but had differing opinions regarding the ranking order of purposes three through five. Almost 39.0% of English teachers selected *communicate to parents* (38.9%) as the third ranked purpose for grading, while 27.8% selected it as the fifth ranked. The option chosen by the majority of English teachers as the fourth purpose for grading was *provide incentive to student/motivation* (43.2%); it is interesting to note that this option was also selected by 31.8% as the fifth purpose for grading, which shows that even within the subject area, teachers were split regarding how they ranked the purposes of grading.

Table 7

Frequencies of the Purpose of Grades Ranking by Subject Area: English

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Measurement of Student's Mastery of Academic Content	42 (65.6)	13 (20.3)	7 (10.9)	1 (1.6)	1 (1.6)	64
Provide Feedback to Students	23 (34.8)	34 (51.5)	7 (10.6)	2 (3.0)	0 (0)	66
Communicate to Parents	0 (0)	1 (1.9)	21 (38.9)	15 (27.8)	15 (27.8)	54
Provide Incentive to Student/Motivation	0 (0)	3 (6.8)	8 (18.2)	19 (43.2)	14 (31.8)	44
Measurement of Student's Level of Effort	0 (0)	10 (25.0)	11 (27.5)	8 (20.0)	8 (20.0)	40
Select, Identify, or Group Students	0 (0)	2 (8.3)	9 (37.5)	4 (16.7)	5 (20.8)	24
Measurement of Student's Level of Responsibility	1 (4.5)	1 (4.5)	1 (4.5)	7 (31.8)	8 (36.4)	22
Evaluate School Programs	0 (0)	1 (6.7)	3 (20.0)	6 (40.0)	1 (6.7)	15
Communicate to School Personnel	0 (0)	1 (9.1)	1 (9.1)	2 (18.2)	2 (18.2)	11

[b] Percentages calculated based on number of respondents selecting each purpose of grading.

[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

As reported in Table 8, the majority of mathematics teachers selected the same options as the top two purposes of grades: *measurement of student's mastery of academic content* (64.7%) and *provide feedback to students* (61.2%) respectively. The purpose selected by the majority of mathematics teachers as the third ranked purpose for grading was *communicate to parents* (34.1%); however, it was also selected by similar percentages of mathematics teachers (19.5%) as number four out of the top five, followed closely by *measurement of student's level of effort* (34.4%). *Provide incentive to student/motivation* was chosen by 38.9% of mathematics teachers as the fifth-ranked purpose for grading. This overlap in ranking choices is consistent with the ranking order that resulted when choices were considered collectively across all subject areas.

Table 8

Frequencies of the Purpose of Grades Ranking by Subject Area: Mathematics

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Measurement of Student's Mastery of Academic Content	33 (64.7)	12 (23.5)	2 (3.9)	3 (5.9)	1 (2.0)	51
Provide Feedback to Students	11 (22.4)	30 (61.2)	7 (14.3)	0 (0)	1 (2.0)	49
Communicate to Parents	3 (7.3)	0 (0)	12 (29.3)	14 (34.1)	8 (19.5)	41
Measurement of Student's Level of Effort	2 (6.3)	4 (12.5)	6 (18.8)	11 (34.4)	7 (21.9)	32
Provide Incentive to Student/Motivation	0 (0)	2 (5.6)	9 (25.0)	5 (13.9)	14 (38.9)	36
Select, Identify, or Group Students	2 (6.1)	3 (9.1)	11 (33.3)	5 (15.2)	5 (15.2)	33
Measurement of Student's Level of Responsibility	0 (0)	1 (5.3)	3 (15.8)	6 (31.6)	5 (26.3)	19
Communicate to School Personnel	0 (0)	0 (0)	1 (5.3)	3 (15.8)	7 (36.8)	19
Evaluate School Programs	1 (7.1)	0 (0)	0 (0)	2 (14.3)	1 (7.1)	14

[b] Percentages calculated based on number of respondents selecting each purpose of grading.

[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Science teachers' ranking choices for the purposes of grades are shown in Table 9. Despite being in agreement on the top two purposes of grading, *measurement of student's mastery of academic content* ranked first with 58.3% of respondents choosing it, and *provide feedback to students* ranked second with 52.2% of respondents choosing it. The third, fourth, and fifth purposes for grading were less clearly defined. The options chosen for rankings three through five had similar percentages of teachers selecting them but at different levels. Thirty-five percent of science teachers chose *measurement of student's level of effort* as the third-ranked purpose; however, 26.3% selected *communicate to parents* and 25.0% ranked *provide incentive to student/motivation* as the third highest purpose for grading. These same three options were also the top choices among science teachers for ranking number four (41.2% for *measurement of student's level of effort*, 37.5% for *provide incentive to student/motivation*, and 26.3% for *communicate to parents*). Science teachers were the only group to rank *measurement of student's level of responsibility* in the top five purposes; 42.9% of science teachers responding selected it at rank number five, but it was closely followed by *communicate to parents* (26.3%) as the fifth-ranked purpose.

Table 9

Frequencies of the Purpose of Grades Ranking by Subject Area: Science

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Measurement of Student's Mastery of Academic Content	14 (58.3)	7 (29.2)	1 (4.2)	1 (4.2)	0 (0)	24
Provide Feedback to Students	8 (34.8)	12 (52.2)	3 (13.0)	0 (0)	0 (0)	23
Measurement of Student's Level of Effort	1 (5.9)	1 (5.9)	6 (35.3)	7 (41.2)	1 (5.9)	17
Provide Incentive to Student/Motivation	0 (0)	1 (6.3)	4 (25.0)	6 (37.5)	4 (25.0)	16
Measurement of Student's Level of Responsibility	0 (0)	2 (14.3)	2 (14.3)	2 (14.3)	6 (42.9)	14
Communicate to Parents	1 (5.3)	0 (0)	5 (26.3)	5 (26.3)	5 (26.3)	19
Evaluate School Programs	0 (0)	0 (0)	0 (0)	2 (20.0)	4 (40.0)	10
Select, Identify, or Group Students	0 (0)	1 (11.1)	3 (33.3)	0 (0)	1 (11.1)	9
Communicate to School Personnel	0 (0)	0 (0)	0 (0)	1 (14.3)	2 (28.6)	7

[b] Percentages calculated based on number of respondents selecting each purpose of grading.

[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Social studies teachers were in agreement regarding the top three purposes of grades. Table 10 shows that the results of social studies teachers choices, similar to the results of all teachers, were split when ranking purposes four and five. Although 38.1% selected *provide incentive to student/motivation* as the fourth purpose, 33.3% selected *measurement of student's level of effort,* and 30.0% selected *communicate to school personnel*. *Select, identify, or group students* was selected by more teachers than other purposes (46.2%) for the fifth ranking.

Table 10

Frequencies of the Purpose of Grades Ranking by Subject Area: Social Studies

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Measurement of Student's Mastery of Academic Content	18 (78.3)	2 (8.7)	1 (4.3)	2 (8.7)	0 (0)	23
Provide Feedback to Students	6 (24.0)	13 (52.0)	2 (8.0)	2 (8.0)	2 (8.0)	25
Communicate to Parents	0 (0)	1 (5.0)	9 (45.0)	3 (15.0)	4 (20.0)	20
Provide Incentive to Student/Motivation	0 (0)	2 (9.5)	4 (19.0)	8 (38.1)	3 (14.3)	21
Measurement of Student's Level of Effort	0 (0)	2 (13.3)	4 (26.7)	5 (33.3)	2 (13.3)	15
Measurement of Student's Level of Responsibility	1 (7.7)	3 (23.1)	3 (23.1)	2 (15.4)	2 (15.4)	13
Select, Identify, or Group Students	0 (0)	1 (7.7)	2 (15.4)	0 (0)	6 (46.2)	13
Communicate to School Personnel	0 (0)	0 (0)	0 (0)	3 (30.0)	3 (30.0)	10
Evaluate School Programs	0 (0)	1 (11.1)	0 (0)	0 (0)	3 (33.3)	9

[b] Percentages calculated based on number of respondents selecting each purpose of grading.

[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Results by Grade Level. Tables 11, 12 and 13 show the survey results for grades six, seven, and eight, respectively. The purposes have been reported by rank order. The percentages were calculated based on the total number of teachers selecting each option, not the total number of respondents.

Respondents were in agreement across grade levels with regard to the top five purposes of grades. The ranking order for purposes one through three was unanimous across the three grade levels and consistent with the top three rankings according to subject area. *Measurement of student's mastery of academic content* was selected number one by 56.3% of sixth grade teachers, 66.7% of seventh grade, and 72.2% of eighth grade respondents. *Provide feedback to students* ranked number two (53.2%, 52.5%, and 57.9% respectively), and *communicate to parents* was ranked third (36.4%, 30.4%, and 39.5% respectively).

Consistent with the overall results and subject area results, there was not a clear preference among teachers for purposes four and five; however, the five options most selected – despite differences in ranking order – were consistent among grade levels. Sixth-grade teachers, as shown in Table 11, were split between two choices for purpose four; 36.7% selected *provide incentive to student/motivation*, while 35.3% selected *measurement of student's level of responsibility*. Interestingly, the same percentage of sixth-grade teachers (35.3%) also selected *measurement of student's level of responsibility* as the fifth purpose of grades, which reinforces the idea that teachers generally agreed on the primary purposes of grading but not necessarily on the order in which to rank them.

Table 11

Frequencies of the Purpose of Grades Ranking by Grade Level: 6th

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Measurement of Student's Mastery of Academic Content	27 (56.3)	13 (27.1)	6 (12.5)	2 (4.2)	0 (0)	48
Provide Feedback to Students	15 (31.9)	25 (53.2)	4 (8.5)	2 (4.3)	1 (2.1)	47
Communicate to Parents	4 (9.1)	1 (2.3)	16 (36.4)	8 (18.2)	11 (25.0)	44
Provide Incentive to Student/Motivation	0 (0)	2 (6.7)	5 (16.7)	11 (36.7)	9 (30.0)	30
Measurement of Student's Level of Effort	2 (7.1)	3 (10.7)	7 (25.0)	7 (25.0)	5 (17.9)	28
Select, Identify, or Group Students	1 (4.0)	2 (8.0)	9 (36.0)	4 (16.0)	6 (24.0)	25
Measurement of Student's Level of Responsibility	0 (0)	2 (11.8)	0 (0)	6 (35.3)	6 (35.3)	17
Communicate to School Personnel	0 (0)	1 (7.7)	1 (7.7)	3 (23.1)	3 (23.1)	13
Evaluate School Programs	0 (0)	0 (0)	0 (0)	3 (23.1)	4 (30.8)	13

[b] Percentages calculated based on number of respondents selecting each purpose of grading.

[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Although it was selected overall as third by seventh-grade teachers, most of them (39.1%) chose *communicate to parents* as the fourth purpose of grades, as shown in Table 12. Tied for fifth place among seventh-grade teachers as a primary purpose of grades were *provide incentive to student/motivation* and *communicate to school personnel*, both selected by a third of the respondents.

Table 12

Frequencies of the Purpose of Grades Ranking by Grade Level: 7th

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Measurement of Student's Mastery of Academic Content	40 (66.7)	13 (21.7)	3 (5.0)	3 (5.0)	1 (1.7)	60
Provide Feedback to Students	17 (28.8)	31 (52.5)	10 (16.9)	1 (1.7)	0 (0)	59
Communicate to Parents	0 (0)	0 (0)	14 (30.4)	18 (39.1)	11 (23.9)	46
Provide Incentive to Student/Motivation	0 (0)	3 (6.7)	11 (24.4)	12 (26.7)	15 (33.3)	45
Measurement of Student's Level of Effort	1 (2.5)	8 (20.0)	8 (20.0)	13 (32.5)	7 (17.5)	40
Select, Identify, or Group Students	1 (3.3)	3 (10.0)	8 (26.7)	2 (6.7)	9 (30.0)	30
Measurement of Student's Level of Responsibility	1 (4.0)	2 (8.0)	6 (24.0)	4 (16.0)	6 (24.0)	25
Communicate to School Personnel	0 (0)	0 (0)	1 (6.7)	3 (20.0)	5 (33.3)	15
Evaluate School Programs	1 (6.3)	1 (6.3)	0 (0)	3 (18.8)	4 (25.0)	16

[b] Percentages calculated based on number of respondents selecting each purpose of grading.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Unlike other grade levels, the rankings made by eighth-grade teachers were clearly delineated; however, as with the other grade levels, there was no clear purpose selected by the majority of respondents for rankings three, four, or five (see Table 13). Although *communicate to parents* was selected as the third purpose (39.5%), it was closely followed by *measurement of student's level of effort* (32.4%) and *select, identify, or group students* (32.0%).

Table 13

Frequencies of the Purpose of Grades Ranking by Grade Level: 8th

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Measurement of Student's Mastery of Academic Content	39 (72.2)	8 (14.8)	3 (5.6)	2 (3.7)	1 (1.9)	54
Provide Feedback to Students	17 (29.8)	33 (57.9)	4 (7.0)	1 (1.8)	2 (3.5)	57
Communicate to Parents	0 (0)	1 (2.3)	17 (39.5)	10 (23.3)	10 (23.3)	43
Provide Incentive to Student/Motivation	0 (0)	2 (4.8)	9 (21.4)	16 (38.1)	11 (26.2)	42
Measurement of Student's Level of Effort	0 (0)	7 (18.9)	12 (32.4)	11 (29.7)	6 (16.2)	37
Measurement of Student's Level of Responsibility	1 (4.0)	3 (12.0)	3 (12.0)	7 (28.0)	8 (32.0)	25
Select, Identify, or Group Students	0 (0)	2 (8.0)	8 (32.0)	3 (12.0)	3 (12.0)	25
Communicate to School Personnel	0 (0)	0 (0)	0 (0)	3 (15.8)	6 (31.6)	19
Evaluate School Programs	0 (0)	1 (5.3)	0 (0)	1 (5.3)	6 (31.6)	19

[b] Percentages calculated based on number of respondents selecting each purpose of grading.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Results by Ability Level. There are three ability levels referenced in this study. Comprehensive refers to classes comprised of average students with mixed academic abilities; honors/gifted classes contain students whose academic ability is above that of average students and therefore are taught at a more rigorous level; and

inclusive/collaborative refers to classes with a mixture of average ability students and students receiving special education services.

When analyzed by student ability level, the same patterns were evident. Tables 14-16 list the rankings by ability level. Like previous tables, the tables have the purposes listed in rank order. The percentages were calculated based on the total number of teachers selecting each option, not the total number of respondents.

Similar to rankings by subject area and grade level, as well as overall results, when analyzed by ability level, purposes one (*measurement of student's mastery of academic content*) and two (*provide feedback to students*) were selected as such by the majority of teachers; differences appeared to some extent at ranking three but more so with rankings four and five.

Table 14 shows the rankings made by teachers of classes with average students, or classes with a comprehensive ability level. Unlike results by subject area and grade level which showed *communicate to parents* consistently ranked as the third purpose, comprehensive teachers chose *select, identify, or group students* as number three (45.5%). Also different from previously described results, this is the first group to choose *evaluate school programs* as a primary purpose, ranked fifth by 42.1% of respondents.

Table 14

Frequencies of the Purpose of Grades Ranking by Ability Level: Comprehensive

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Measurement of Student's Mastery of Academic Content	56 (68.3)	16 (19.5)	8 (9.8)	1 (1.2)	1 (1.2)	82
Provide Feedback to Students	21 (25.6)	49 (59.8)	8 (9.8)	3 (3.7)	1 (1.2)	82
Communicate to Parents	3 (4.5)	0 (0)	24 (35.8)	19 (28.4)	17 (19.5)	67
Provide Incentive to Student/Motivation	0 (0)	4 (6.6)	8 (13.1)	23 (37.7)	23 (37.7)	61
Measurement of Student's Level of Effort	2 (4.3)	7 (14.9)	16 (34.0)	13 (27.7)	6 (12.8)	47
Select, Identify, or Group Students	1 (3.0)	2 (6.1)	15 (45.5)	5 (15.2)	8 (24.2)	33
Measurement of Student's Level of Responsibility	1 (3.8)	5 (19.2)	2 (7.7)	8 (30.8)	8 (30.8)	26
Communicate to School Personnel	0 (0)	1 (4.5)	2 (9.1)	4 (18.2)	9 (40.9)	22
Evaluate School Programs	0 (0)	0 (0)	0 (0)	6 (31.6)	8 (42.1)	19

[b] Percentages calculated based on number of respondents selecting each purpose of grading.

[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Teachers referencing classes with honors and/or gifted students had rankings, shown in Table 15, that were in line with overall, subject area, and grade-level rankings. The top five purposes were chosen from the same five categories, with the majority of respondents choosing purposes one (*measurement of student's mastery of academic content* – 64.4%) and two (*provide feedback to students* – 52.3%).

Table 15

Frequencies of the Purpose of Grades Ranking by Ability Level: Honors and/or Gifted

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Measurement of Student's Mastery of Academic Content	42 (64.4)	13 (20.0)	3 (4.6)	6 (9.2)	0 (0)	65
Provide Feedback to Students	22 (33.8)	34 (52.3)	7 (10.8)	1 (1.5)	1 (1.5)	65
Communicate to Parents	1 (1.8)	2 (3.5)	21 (36.8)	13 (22.8)	15 (26.3)	57
Measurement of Student's Level of Effort	1 (2.0)	9 (18.4)	10 (20.4)	14 (28.6)	11 (22.4)	49
Provide Incentive to Student/Motivation	0 (0)	2 (4.5)	14 (31.8)	13 (29.5)	7 (15.9)	44
Measurement of Student's Level of Responsibility	0 (0)	2 (5.7)	5 (14.3)	9 (25.7)	10 (28.6)	35
Select, Identify, or Group Students	1 (2.9)	4 (11.8)	6 (17.6)	2 (5.9)	7 (20.6)	34
Communicate to School Personnel	0 (0)	0 (0)	0 (0)	4 (18.2)	5 (22.7)	22
Evaluate School Programs	0 (0)	1 (4.8)	0 (0)	0 (0)	4 (19.0)	21

[b] Percentages calculated based on number of respondents selecting each purpose of grading.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Table 16 shows rankings from those teachers who referenced an inclusive or collaborative class. These classes contain average-level students as well as special education students. The sample size of inclusive/collaborative teachers is small (n = 17), so these results may not generalize to the broader teacher population at the district level.

Purpose two overlapped with purpose one for this group: the majority (53.3%) selected *measurement of student's mastery of academic content* as the primary purpose, but it was also chosen by 33.3% of respondents as the second purpose. *Provide feedback to students* was the second choice for the number two ranked purpose of grades, as it was selected by 31.3% of respondents for that ranking. It is interesting to note that, of the ability-level responses, this was the only group to rank *measurement of student's level of responsibility* as a top five purpose of grades; almost 43.0% of comprehensive/ collaborative teachers chose it as the fifth purpose.

Table 16

Frequencies of the Purpose of Grades Ranking by Ability Level: Inclusive/Collaborative

Purposes of Grades	Rank #1 *n* (%)	Rank #2 *n* (%)	Rank #3 *n* (%)	Rank #4 *n* (%)	Rank #5 *n* (%)[b]	Total *n*[c]
Provide Feedback to Students	6 (37.5)	5 (31.3)	4 (25.0)	0 (0)	1 (6.3)	16
Measurement of Student's Mastery of Academic Content	8 (53.3)	5 (33.3)	1 (6.7)	0 (0)	1 (6.7)	15
Select, Identify, or Group Students	0 (0)	1 (8.3)	4 (33.3)	2 (16.7)	2 (16.7)	12
Provide Incentive to Student/Motivation	0 (0)	2 (16.7)	2 (16.7)	3 (25.0)	5 (41.7)	12
Communicate to Parents	0 (0)	0 (0)	2 (20.0)	5 (50.0)	0 (0)	10
Measurement of Student's Level of Effort	0 (0)	2 (25.0)	1 (12.5)	3 (37.5)	1 (12.5)	8
Evaluate School Programs	1 (12.5)	1 (12.5)	0 (0)	1 (12.5)	2 (25.0)	8
Measurement of Student's Level of Responsibility	1 (14.3)	0 (0)	2 (28.6)	0 (0)	3 (42.9)	7
Communicate to School Personnel	0 (0)	0 (0)	0 (0)	1 (33.3)	0 (0)	3

[b] Percentages calculated based on number of respondents selecting each purpose of grading.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Attitudes Toward Grading

Several survey questions were used to examine teachers' attitudes towards different aspects of grading. Measurement experts assert that grades should communicate one message: how well the student has mastered the content standard and/or performance objective (Ebel & Frisbie, 1991; Guskey, 1994; Lambating & Allen, 2002). Teachers were

presented with 14 statements designed to measure their attitudes using a likert-style agreement scale that ranged from strongly disagree to strongly agree. Twelve of the statements referenced grading practices that measurement experts indicated should be avoided when determining grades and therefore grouped, for this study, into the category *non-academic factors*. Two questions addressed factors that measurement experts recommend when assigning grades and were grouped as *academic factors*: items measurement experts assert can, and should, be communicated by the grades teachers assign.

By categorizing the survey items as non-academic or academic factors, Table 17, which provides the results of those responding to the statements, shows that teachers' attitudes toward what should factor into a grade were not well aligned with the recommendations of measurement experts. For example, 92.0% agreed or strongly agreed that grades should reflect progress/improvement, and 81.2% agreed or strongly agreed that grades should reflect student's effort. Both of these statements were categorized as *non-academic factors* because they are contrary to grading factors recommended by measurement experts (Guskey, 1994). These results are similar to previous studies that found a lack of alignment between classroom practices and theoretical recommendations (Brookhart, 1993; Cross & Frary, 1999). Large numbers of the teachers participating in this study either agreed or strongly agreed with statements that allow the inclusion of non-academic factors in the determination of grades. However, they also overwhelming agreed with the academic factors: 95.1% agreed or strongly agreed that grades should reflect

students' mastery of the content, and over half agreed that criterion-referenced grading is a legitimate method of identifying a student's ability.

Table 17

Percentage of Teachers Selecting "Strongly Agree" or "Agree" to Statements on Grading Practices

Grading Practices	*n*	%
Non-Academic Factors		
Grades should reflect progress/improvement	149	92.0
Grades should reflect student's effort	129	81.2
Homework completion should be included in final grades	127	77.9
Grades should be used to motivate students	116	71.2
All assignments should count towards the course grade	106	65.0
Grades should reflect student's ability to follow directions	88	54.3
Grades should be used to teach students responsibility	77	47.3
Students should receive zeros for incomplete work	72	44.2
Grades should be used to reward students	60	36.8
Homework accuracy should be included in final grades	52	31.9
Norm-referenced grading is a legitimate method of identifying a student's ability	48	30.2
All assignments should be graded	23	14.1
Academic Factors		
Grades should reflect student's mastery of the content	155	95.1
Criterion-referenced grading is a legitimate method of identifying a student's ability	93	58.5

Results by Subject Area. When analyzing the results by subject area, results show that teachers across all subjects agreed or strongly agreed with practices in both categories. As shown in Table 18, the strongest agreement among teachers in all content areas was with the statement, under the category *academic factors,* that grades should reflect mastery: science (100.0%), social studies (100.0%), English (93.7%), and mathematics (92.2%). Contrary to recommendations by measurement experts that grades should not incorporate improvement over time (Guskey, 1994), a large percentage of teachers (98.0% mathematics, 95.8% social studies, 87.5% science, and 87.1% English) participating in this study reported agreement to grades reflecting progress/improvement.

Agreement with other statements regarding grading practices varied among subjects, regardless of whether the statement was related to academic or non-academic factors. There was strong agreement among science teachers that grades should reflect how much effort students display (95.7%); English and social studies teachers also agreed (83.3% respectively). Over 80.0% of science (83.3%), social studies (83.3%), and mathematics (82.4%) teachers agreed that homework completion should be incorporated into end-of-course grades.

Although there was strong agreement across subject areas, there were some differences, especially in the responses of mathematics teachers, in the percentages of those who selected "agree" or "strongly agree" to certain grading practices. Whereas almost 80.0% of science and social studies teachers and almost 64.0% of English teachers agreed or strongly agreed that all assignments should count towards the course grade, only 52.9% of mathematics teachers responded the same way. Greater than 60.0% of English,

science, and social studies teachers agreed or strongly agreed that criterion-referenced grading is a legitimate method of identifying a student's ability, contrasted to 47.9% of mathematics teachers. Mathematics teachers were also less likely to agree or strongly agree that homework accuracy should be included in final grades: (15.7% mathematics, 29.2% science, 38.1% English, and 50.0% social studies). English teachers differed in their agreement with the other subjects regarding the statement that homework completion should be included in final grades; greater than 80.0% of mathematics, science and social studies teachers agreed or strongly agreed, compared to 69.8% of English teachers.

Table 18

Percentage of Teachers Selecting "Strongly Agree" or "Agree" to Statements on Grading Practices: Subject Area Results

Grading Practices	*English* *n (%)*	*Mathematics* *n (%)*	*Science* *n (%)*	*Social Studies* *n (%)*[c]
Non-Academic Factors				
Grades should reflect progress/improvement	54 (87.1)	50 (98.0)	21 (87.5)	23 (95.8)
Grades should reflect student's effort	50 (83.3)	37 (72.5)	22 (95.7)	20 (83.3)
Homework completion should be included in final grades	44 (69.8)	42 (82.4)	20 (83.3)	20 (83.3)
Grades should be used to motivate students	48 (76.2)	31 (60.8)	15 (62.5)	21 (87.5)
All assignments should count towards the course grade	40 (63.5)	27 (52.9)	19 (79.2)	19 (79.2)
Grades should reflect student's ability to follow directions	32 (51.6)	24 (47.1)	17 (70.8)	15 (62.5)
Grades should be used to teach students responsibility	24 (38.1)	25 (49.0)	16 (66.7)	12 (50.0)
Students should receive zeros for incomplete work	22 (34.9)	24 (47.1)	12 (50.0)	14 (58.3)
Grades should be used to reward students	21 (33.3)	16 (31.4)	11 (45.8)	12 (50.0)
Homework accuracy should be included in final grades	24 (38.1)	8 (15.7)	7 (29.2)	12 (50.0)
Norm-referenced grading is a legitimate method of identifying a student's ability	19 (30.6)	14 (29.2)	6 (25.0)	9 (37.5)
All assignments should be graded	8 (12.7)	6 (11.8)	6 (25.0)	3 (12.5)
Academic Factors				
Grades should reflect student's mastery of the content	59 (93.7)	47 (92.2)	24 (100.0)	24 (100.0)
Criterion-referenced grading is a legitimate method of identifying a student's ability	38 (61.3)	23 (47.9)	15 (62.5)	17 (70.8)

[c] Total *n* count by subject area: English (68), Mathematics (54), Science (25), Social Studies (27)

Results by Grade Level. Table 19 shows that grade-level results echoed those by subject area. Greater than 90.0% of teachers responding agreed that grades should reflect mastery of the content (91.8% of 6th grade, 94.8% of 7th grade, and 98.2% of 8th grade teachers). Content mastery was followed closely by grades reflecting progress/improvement: 89.8%, 93.0%, and 92.7% at grades 6, 7, and 8 respectively. Based on their agreement, seventh- and eighth-grade teachers (80.7% and 84.6%) were more likely to include student effort when assigning grades than are their 6th grade counterparts. Also consistent among all grade levels is the fact that greater than three quarters of teachers agreed that homework completion should be included in final grades.

The responses of sixth grade teachers, however, showed that they are less likely than teachers in seventh grade and eighth grade to agree with statements that indicate grades should motivate students (59.2% of 6th grade vs. 70.7% of 7th grade and 81.8% of 8th grade) and teach responsibility (36.7%, 50.0%, and 52.7% respectively). Sixth-grade teachers also indicated that they do not grade every assignment (8.2%), nor do they believe homework accuracy should be included (16.3%).

Table 19

Percentage of Teachers Selecting "Strongly Agree" or "Agree" to Statements on Grading Practices: Grade Level Results

Grading Practices	6^{th} *n* (%)	7^{th} *n* (%)	8^{th} *n* (%)[c]
Non-Academic Factors			
Grades should reflect progress/improvement	44 (89.8)	53 (93.0)	51 (92.7)
Grades should reflect student's effort	38 (77.6)	46 (80.7)	44 (84.6)
Homework completion should be included in final grades	37 (75.5)	46 (79.3)	43 (78.2)
Grades should be used to motivate students	29 (59.2)	41 (70.7)	45 (81.8)
All assignments should count towards the course grade	32 (65.3)	37 (63.8)	36 (65.5)
Grades should reflect student's ability to follow directions	23 (46.9)	30 (52.6)	34 (61.8)
Grades should be used to teach students responsibility	18 (36.7)	29 (50.0)	29 (52.7)
Students should receive zeros for incomplete work	24 (49.0)	24 (41.4)	23 (41.8)
Grades should be used to reward students	17 (34.7)	18 (31.0)	24 (43.6)
Homework accuracy should be included in final grades	8 (16.3)	23 (39.7)	21 (38.2)
Norm-referenced grading is a legitimate method of identifying a student's ability	15 (31.3)	20 (35.7)	13 (24.1)
All assignments should be graded	4 (8.2)	8 (13.8)	11 (20.0)
Academic Factors			
Grades should reflect student's mastery of the content	45 (91.8)	55 (94.8)	54 (98.2)
Criterion-referenced grading is a legitimate method of identifying a student's ability	30 (62.5)	37 (66.1)	26 (48.1)

[c] Total *n* count by grade level: 6^{th} grade (51), 7^{th} grade (64), 8^{th} grade (59)

Results by Ability Level. Teachers of students in an inclusive/collaborative setting responded differently from those in comprehensive and honors/gifted classes, with more agreeing with statements about grading practices that are less directly reflective of students' academic knowledge (e.g., homework completion as opposed to correctness and student effort). As shown in Table 20, a smaller percentage of inclusive/collaborative teachers agreed or strongly agreed that grades should reflect mastery (87.5% as compared to 96.8% of honors/gifted and 95.2% of comprehensive teachers). Over 90.0% of comprehensive teachers (92.7%) and honors/gifted teachers (95.2%) agreed that grades should reflect progress/improvement, but only 75.0% of inclusive/collaborative teachers indicated agreement. Statements that yielded a greater percentage of agreement from teachers in inclusive/collaborative settings were those that were indirectly related to students' content knowledge and skills: homework completion (87.5%), effort (81.3%), and grades as a motivator of students (75.0%). Teachers of honors/gifted students tended to agree with having grades reflect students' efforts (83.6%) at a rate higher than teachers in an inclusive/collaborative class. Homework completion is included, based on the agreement indicated, by 76.2% of teachers of honors/gifted classes, and 74.1% agreed that grades should be used to motivate students. By contrast, approximately three-fourths of comprehensive teachers agreed that homework completion (77.1%) and effort (79.0%) should be included, and less than 70.0% agreed with using grades to motivate students.

Table 20

Percentage of Teachers Selecting "Strongly Agree" or "Agree" to Statements on Grading Practices: Ability Level Results

Grading Practices	*Comprehensive* *n* (%)	*Honors/Gifted* *n* (%)	*Inclusive/Collaborative* *n* (%)
Non-Academic Factors			
Grades should reflect progress/improvement	76 (92.7)	60 (95.2)	12 (75.0)
Grades should reflect student's effort	64 (79.0)	51 (83.6)	13 (81.3)
Homework completion should be included in final grades	64 (77.1)	48 (76.2)	14 (87.5)
Grades should be used to motivate students	58 (69.9)	45 (71.4)	12 (75.0)
All assignments should count towards the course grade	57 (68.7)	39 (61.9)	9 (56.3)
Grades should reflect student's ability to follow directions	42 (51.2)	38 (60.3)	7 (43.8)
Grades should be used to teach students responsibility	36 (43.3)	33 (52.4)	7 (43.8)
Students should receive zeros for incomplete work	33 (39.8)	33 (52.4)	5 (31.3)
Grades should be used to reward students	28 (33.7)	25 (39.7)	6 (37.5)
Homework accuracy should be included in final grades	33 (39.8)	16 (25.4)	2 (12.5)
Norm-referenced grading is a legitimate method of identifying a student's ability	28 (34.1)	18 (29.5)	2 (13.3)
All assignments should be graded	13 (15.7)	9 (14.3)	1 (6.3)
Academic Factors			
Grades should reflect student's mastery of the content	79 (95.2)	61 (96.8)	14 (87.5)
Criterion-referenced grading is a legitimate method of identifying a student's ability	51 (62.2)	36 (59.0)	6 (40.0)

[c] Total *n* count by ability level: Comprehensive (87), Honors/Gifted (70), Inclusive/Collaborative (17)

Assessment and Grading Practices

The second research question that guided this study asked what assessment and grading practices do middle school, core academic subject teachers use most often in their classrooms when determining students' grades. Survey participants were asked to respond to several questions designed to determine how often teachers used certain types of assessments and how much influence specific grading practices had on students' end-of-course grades. Survey questions for this section were grouped, based on the grading and assessment literature, into three categories: academic factors, non-academic factors, and assessment methods.

Overall Results. Similar to the survey question that asked teachers to rank the purpose of grades, respondents were also asked to rank items focused on assessment methods. Teachers were asked to indicate how much influence certain practices have on grading, and they were asked to rank what they believed to be the top five assessment practices from a list containing 15 options. Results are listed in Table 21 in order of the total number of respondents that choose the assessment practice, rather than in rank order. Percentages were calculated based on the number of teachers selecting the option, rather than the total number of respondents. The results displayed in the tables show that the assessment method chosen by the greater number of teachers are not necessarily those ranked with the largest percentage of respondents.

Table 21 shows the assessment practices in order from most to least selected. Overall, *teacher-developed assessments* was chosen by more respondents (126) as a top five method of assessment, followed by *grades* (105), *teacher questioning and*

observations (103), *portfolios* (83), and *pre/post-tests* (79). When determining what assessment practices ranked in the top five, however, differences were evident in the order and in the selections.

The number one-ranked assessment method, with 38.6% of respondents selecting it first, was *portfolios*; number two was *teacher questioning and observations* (29.1%); and third was *grades* (26.7%). Each of these practices was also among the five most selected; however, the fourth- and fifth-ranked practices were not among the five chosen by the greater number of teachers. *Self-assessments* by students was selected by 25.5% as fourth ranked, yet it was selected by only 47 respondents; the fifth-ranked assessment practice, *state assessments*, was chosen among the top five by 52 teachers – placing it seventh out of the 15 options – but with 36.5% choosing it fifth, it out-ranked other practices that were selected by higher numbers of respondents.

Table 21

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
Teacher-developed assessments (*n* = 126)	43 (34.1)	31 (24.6)	21 (16.7)	15 (11.9)	15 (11.9)
Grades on assignments (*n* = 105)	23 (21.9)	23 (21.9)	28 (26.7)	15 (14.3)	11 (10.5)
Teacher questioning and observations (*n* = 103)	14 (13.6)	30 (29.1)	19 (18.4)	21 (20.4)	15 (14.6)
Portfolios (*n* = 83)	32 (38.6)	12 (14.5)	12 (14.5)	9 (10.8)	8 (9.6)
Pre/post-tests (unit) (*n* = 79)	22 (27.8)	14 (17.7)	13 (16.5)	15 (19.0)	4 (5.1)
Projects/reports (*n* = 78)	8 (10.3)	19 (24.4)	11 (14.1)	14 (17.9)	16 (20.5)
State assessments (SOLs) (*n* = 52)	1 (1.9)	6 (11.5)	3 (5.8)	9 (17.3)	19 (36.5)
Class participation (*n* = 52)	6 (11.5)	7 (13.5)	11 (21.2)	10 (19.2)	9 (17.3)
District assessments/ benchmarks tests (*n* = 51)	4 (7.8)	5 (9.8)	5 (9.8)	9 (17.6)	14 (27.5)
End of lesson review (independent) (*n* = 49)	5 (10.2)	4 (8.2)	6 (12.2)	10 (20.4)	11 (22.4)
Self-assessments (students') (*n* = 47)	1 (2.1)	3 (6.4)	12 (25.5)	12 (25.5)	8 (17.0)
Homework completion/ quality (*n* = 36)	1 (2.8)	2 (5.6)	9 (25.0)	5 (13.9)	8 (22.2)
End of lesson review (teacher-led) (*n* = 35)	0 (0)	2 (5.7)	8 (22.9)	5 (14.3)	9 (25.7)
Group work (*n* = 31)	0 (0)	0 (0)	1 (3.2)	6 (19.4)	8 (25.8)
Behavior/attitude in class (*n* = 25)	0 (0)	2 (8.0)	1 (4.0)	4 (16.0)	3 (12.0)

[b] Percentages calculated based on number of respondents selecting each assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Responses to questions regarding the amount of influence certain factors had on the grade assigned have been grouped into three categories: academic factors, non-academic factors, and questions/assessment methods.

Academic Factors. Several survey questions asked teachers to indicate how much influence various factors had on the end-of-course grades they assigned to students. The responses are reported in Table 22. Mastery of objectives (95.1%), quiz scores (96.8%), and test scores (98.7%) held the most influence, with the highest collective percentages of respondents selecting some, quite a bit, and extensive influence. Respondents indicated they allowed student improvement to have some influence (46.3%) or quite a bit of influence (22.2%). The inclusion of zeros as an influence on students' grades was split among respondents, with 41.4% choosing minimum to no influence at all and 53.7% selecting some or quite a bit. Classwork and quiz scores had very few respondents, 5.5% and 3.1% respectively, indicating these assessments carried minimum or no influence at all on students' end-of-course grades. Classwork and quiz scores were considered by the majority of teachers when deciding which grades to give to students. One-fifth of teachers reported that daily homework was not influential when determining grades. Academic extra credit had no influence for 19.6% of respondents, minimal influence for 42.9%, and some influence for 36.2% of respondents.

Table 22

Academic Factors for Determining End-of-Course Grades

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	1 (0.6)	8 (4.9)	58 (35.6)	72 (44.2)	24 (14.7)
Daily homework	4 (2.5)	31 (19.0)	71 (43.6)	45 (27.6)	12 (7.4)
Extra credit – academic	32 (19.6)	70 (42.9)	59 (36.2)	2 (1.2)	0 (0)
Improvement of performance	23 (14.2)	24 (14.8)	75 (46.3)	36 (22.2)	4 (2.5)
Inclusion of zeros	26 (16.0)	41 (25.3)	56 (34.6)	31 (19.1)	8 (4.9)
Mastery of specific learning objectives	2 (1.2)	6 (3.7)	29 (17.9)	78 (48.1)	47 (29.0)
Performance compared with peers	80 (49.7)	41 (25.5)	30 (18.6)	8 (5.0)	2 (1.2)
Projects	4 (2.5)	16 (9.8)	54 (33.1)	67 (41.1)	22 (13.5)
Quiz scores	1 (0.6)	4 (2.5)	31 (19.1)	95 (58.6)	31 (19.1)
Test scores	1 (0.6)	1 (0.6)	25 (15.4)	83 (51.2)	52 (32.1)

Non-Academic Factors. Items measurement experts recommend that teachers exclude when determining students' grades have been grouped together in this study as non-academic factors. As reported in Table 23, non-academic factors selected as having the most influence on students' grades were student effort (42.3% selected some, 31.3%

quite a bit, and 4.9% extensive) and student participation (some 43.6%, quite a bit 25.2%, and extensive 4.9%) influence. Respondents were split on how much weight work habits had when they determined students grades: 56.1% selected minimum or no influence, while 43.8% selected some, quite a bit, or extensive. They were also split on the influence of the students' behavior/attitude in class: 33.1% selected no influence at all for behavior/attitude in class as a factor when determining end-of-course grades; 48.5% reported minimum/some influence; and 18.4% considered behavior/attitude quite a bit/extensively. Twenty-six percent of respondents indicated that attendance in class had no influence at all on a student's end-of-course grade, while 44.1% reported it had some/minimal influence, and 23.9% ranked attendance as having quite a bit/extensive influence on a student's end-of-course grade. Very few respondents reported that non-academic extra credit had a moderate to strong influence on a student's grade (10.5%).

Table 23

Non-Academic Factors for Determining End-of-Course Grades

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	52 (31.9)	36 (22.1)	36 (22.1)	28 (17.2)	11 (6.7)
Behavior/ attitude in class	54 (33.1)	46 (28.2)	33 (20.2)	22 (13.5)	8 (4.9)
Extra credit – non-academic	91 (56.2)	54 (33.3)	16 (9.9)	1 (0.6)	0 (0)
Student effort	14 (8.6)	21 (12.9)	69 (42.3)	51 (31.3)	8 (4.9)
Participation	15 (9.2)	28 (17.2)	71 (43.6)	41 (25.2)	8 (4.9)
Work habits (neatness, etc.)	39 (24.1)	52 (32.1)	48 (29.6)	18 (11.1)	5 (3.1)

Types of Questions. As documented in Table 24, respondents indicated they frequently used multiple-choice questions (58.4%) and short-answer questions (54.7%); student exhibits were used frequently by 36.9% of respondents and occasionally by 36.3%. Approximately 30.0% of survey participants indicated they used portfolios frequently, compared to 24.5% who reported occasional use, 23.3% who selected seldom, and 15.7 % who never used portfolios. Essays were used occasionally or frequently by 58.8%, while 35.6% of teachers selected seldom or never. The majority of respondents (61.0%) seldom or never used true-false questions; matching questions were used occasionally by 41.3% and frequently by 29.4% of respondents.

Table 24

Frequencies for Types of Questions

Types of Question	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	27 (16.9)	30 (18.8)	44 (27.5)	50 (31.3)	9 (5.6)
Matching	8 (5.0)	31 (19.4)	66 (41.3)	47 (29.4)	8 (5.0)
Multiple Choice	0 (0)	8 (5.0)	37 (23.0)	94 (58.4)	22 (13.7)
Short Answer	3 (1.9)	15 (9.3)	39 (24.2)	88 (54.7)	16 (9.9)
True-False	36 (22.6)	61 (38.4)	39 (24.5)	16 (10.1)	7 (4.4)
Student Exhibits (projects & reports)	5 (3.1)	26 (16.3)	58 (36.3)	59 (36.9)	12 (7.5)
Portfolios	25 (15.7)	37 (23.3)	39 (24.5)	47 (29.6)	11 (6.9)

Results by Subject Area. When analyzed by content area, the selections made by teachers were rather consistent. Based on the total number of respondents who ranked assessment practices, greater numbers of teachers across all subject areas chose *teacher-developed assessments, teacher questioning and observations*, and *grades on assignments* among the top five assessment methods. The ranking order is slightly different from the order in which assessment methods are listed because the order is based on the total number of teachers who selected the item as a top five assessment method rather than by rank order.

Table 25 lists the results of English teachers' responses. Rounding out the top five assessment methods most often selected were *portfolios* and *project/reports*. Yet, when looking at the rank ordering, *portfolios* ranked first (48.7%). *Teacher questioning and observations* and *grades on assignments* ranked second and third; however, those practices that ranked fourth (*end of lesson review)* and fifth (*state assessments*) were not among the five most selected options.

Similarly, three methods that ranked in the top five selected by mathematics teachers differed from the five practices most often selected by the subject area cohort (see Table 26). Teachers of mathematics ranked students' *self-assessments* third, *end of lesson review* fourth, and *state assessments* fifth.

Table 27 shows that s*tate assessments* ranked highest (2^{nd}) for science teachers. Science teachers were divided in their choice for the fifth-ranked assessment practice: both *district assessments/benchmark tests* and *homework completion/quality* were so ranked by 33.3%.

Only one of the most selected assessment practices (*teacher-developed assessments*) was also ranked in the top five by social studies teachers; *teacher-developed assessments* was selected by more teachers than any other practice, and it was ranked number one (Table 28). *Portfolios* ranked second, and teachers split on the third-ranked assessment practice. *Homework completion/quality* and *end of lesson review* were both chosen third by 40.0% of respondents.

Table 25

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods by Subject Area: English

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
Teacher-developed assessments (*n* = 45)	14 (31.1)	13 (28.9)	7 (15.6)	5 (11.1)	5 (11.1)
Teacher questioning and observations (*n* = 43)	6 (14.0)	16 (37.2)	4 (9.3)	8 (18.6)	7 (16.3)
Grades on assignments (*n* = 41)	7 (17.1)	10 (24.4)	14 (34.1)	4 (9.8)	5 (12.2)
Portfolios (*n* = 39)	19 (48.7)	1 (2.6)	5 (12.8)	3 (7.7)	6 (15.4)
Projects/reports (*n* = 35)	3 (8.6)	9 (25.7)	4 (11.4)	8 (22.9)	6 (17.1)
Pre/post-tests (unit) (*n* = 26)	5 (19.2)	4 (15.4)	6 (23.1)	5 (19.2)	2 (7.7)
Class participation (*n* = 23)	3 (13.0)	4 (17.4)	5 (21.7)	4 (17.4)	5 (21.7)
Self-assessments (students') (*n* = 20)	1 (5.0)	0 (0)	4 (20.0)	5 (25.0)	5 (25.0)
State assessments (SOLs) (*n* = 19)	0 (0)	1 (5.3)	2 (10.5)	3 (15.8)	7 (36.8)
End of lesson review (independent) (*n* = 19)	2 (10.5)	0 (0)	4 (21.1)	6 (31.6)	3 (15.8)
District assessments/ benchmarks tests (*n* = 18)	1 (5.6)	2 (11.1)	2 (11.1)	3 (16.7)	4 (22.2)
Group work (*n* = 12)	0 (0)	0 (0)	1 (8.3)	3 (25.0)	2 (16.7)
Homework completion/ quality (*n* = 11)	1 (9.1)	0 (0)	1 (9.1)	3 (27.3)	2 (18.2)
End of lesson review (teacher-led) (*n* = 10)	0 (0)	1 (10.0)	3 (30.0)	0 (0)	3 (30.0)
Behavior/attitude in class (*n* = 8)	0 (0)	1 (12.5)	0 (0)	2 (25.0)	0 (0)

[b] Percentages calculated based on number of respondents selecting the assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Table 26

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods by Subject Area: Mathematics

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
Teacher-developed assessments (*n* = 39)	12 (30.8)	13 (33.3)	5 (12.8)	4 (10.3)	5 (12.8)
Teacher questioning and observations (*n* = 32)	5 (15.6)	12 (37.5)	8 (25.0)	6 (18.8)	0 (0)
Grades on assignments (*n* = 32)	11 (34.4)	6 (18.8)	6 (18.8)	2 (6.3)	4 (12.5)
Pre/post-tests (unit) (*n* = 27)	7 (25.9)	5 (18.5)	4 (14.8)	6 (22.2)	0 (0)
Portfolios (*n* = 21)	6 (28.6)	1 (4.8)	6 (28.6)	3 (14.3)	2 (9.5)
Projects/reports (*n* = 20)	2 (10.0)	4 (20.0)	2 (10.0)	3 (15.0)	6 (30.0)
State assessments (SOLs) (*n* = 20)	1 (5.0)	2 (10.0)	1 (5.0)	4 (20.0)	7 (35.0)
District assessments/ benchmarks tests (*n* = 19)	3 (15.8)	1 (5.3)	1 (5.3)	4 (21.1)	6 (31.6)
End of lesson review (independent) (*n* = 18)	2 (11.1)	1 (5.6)	1 (5.6)	3 (16.7)	6 (33.3)
Class participation (*n* = 17)	2 (11.8)	2 (11.8)	4 (23.5)	4 (23.5)	1 (5.9)
End of lesson review (teacher-led) (*n* = 16)	0 (0)	1 (6.3)	2 (12.5)	5 (31.3)	3 (18.8)
Homework completion/ quality (*n* = 14)	0 (0)	2 (14.3)	5 (35.7)	2 (14.3)	1 (7.1)
Group work (*n* = 13)	0 (0)	0 (0)	0 (0)	3 (23.1)	4 (30.8)
Behavior/attitude in class (*n* = 10)	0 (0)	0 (0)	1 (10.0)	0 (0)	3 (30.0)
Self-assessments (students') (*n* = 11)	0 (0)	1 (9.1)	5 (45.5)	1 (9.1)	1 (9.1)

[b] Percentages calculated based on number of respondents selecting the assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Table 27

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods by Subject Area: Science

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
Teacher-developed assessments (*n* = 21)	8 (38.1)	1 (4.8)	4 (19.0)	4 (19.0)	4 (19.0)
Grades on assignments (*n* = 15)	2 (13.3)	4 (26.7)	3 (20.0)	4 (26.7)	1 (6.7)
Pre/post-tests (unit) (*n* = 15)	7 (46.7)	3 (20.0)	2 (13.3)	0 (0)	1 (6.7)
Teacher questioning and observations (*n* = 14)	1 (7.1)	1 (7.1)	5 (35.7)	3 (21.4)	4 (28.6)
Portfolios (*n* = 13)	5 (38.5)	3 (23.1)	1 (7.7)	3 (23.1)	0 (0)
Self-assessments (students') (*n* = 12)	0 (0)	2 (16.7)	2 (16.7)	5 (41.7)	1 (8.3)
Projects/reports (*n* = 10)	0 (0)	3 (30.0)	2 (20.0)	1 (10.0)	3 (30.0)
Class participation (*n* = 7)	0 (0)	1 (14.3)	1 (14.3)	1 (14.3)	2 (28.6)
End of lesson review (independent) (*n* = 7)	0 (0)	2 (28.6)	0 (0)	0 (0)	2 (28.6)
State assessments (SOLs) (*n* = 6)	0 (0)	2 (33.3)	0 (0)	1 (16.7)	0 (0)
District assessments/ benchmarks tests (*n* = 6)	0 (0)	0 (0)	1 (16.7)	0 (0)	2 (33.3)
Homework completion/ quality (*n* = 6)	0 (0)	0 (0)	1 (16.7)	0 (0)	2 (33.3)
Behavior/attitude in class (*n* = 5)	0 (0)	1 (20.0)	0 (0)	1 (20.0)	0 (0)
End of lesson review (teacher-led) (*n* = 4)	0 (0)	0 (0)	1 (25.0)	0 (0)	1 (25.0)
Group work (*n* = 3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

[b] Percentages calculated based on number of respondents selecting the assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Table 28

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods by Subject Area: Social Studies

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
Teacher-developed assessments (*n* = 20)	9 (45.0)	4 (20.0)	4 (20.0)	2 (10.0)	1 (5.0)
Grades on assignments (*n* = 16)	3 (18.8)	2 (12.5)	5 (31.3)	5 (31.3)	1 (6.3)
Teacher questioning and observations (*n* = 13)	2 (15.4)	1 (7.7)	2 (15.4)	4 (30.8)	3 (23.1)
Projects/reports (*n* = 12)	2 (16.7)	3 (25.0)	3 (25.0)	2 (16.7)	1 (8.3)
Pre/post-tests (unit) (*n* = 11)	3 (27.3)	2 (18.2)	1 (9.1)	4 (36.4)	1 (9.1)
Portfolios (*n* = 10)	2 (20.0)	7 (70.0)	0 (0)	0 (0)	0 (0)
State assessments (SOLs) (*n* = 7)	0 (0)	1 (14.3)	0 (0)	1 (14.3)	5 (71.4)
District assessments/ benchmarks tests (*n* = 7)	0 (0)	2 (28.6)	1 (14.3)	1 (14.3)	2 (28.6)
Homework completion /quality (*n* = 5)	0 (0)	0 (0)	2 (40.0)	0 (0)	3 (60.0)
End of lesson review (teacher-led) (*n* = 5)	0 (0)	0 (0)	2 (40.0)	0 (0)	2 (40.0)
Class participation (*n* = 5)	1 (20.0)	0 (0)	1 (20.0)	1 (20.0)	1 (20.0)
End of lesson review (independent) (*n* = 5)	1 (20.0)	1 (20.0)	1 (20.0)	1 (20.0)	0 (0)
Self-assessments (students') (*n* = 4)	0 (0)	0 (0)	1 (25.0)	1 (25.0)	1 (25.0)
Group work (*n* = 3)	0 (0)	0 (0)	0 (0)	0 (0)	2 (66.7)
Behavior/attitude in class (*n* = 2)	0 (0)	0 (0)	0 (0)	1 (50.0)	0 (0)

[b] Percentages calculated based on number of respondents selecting the assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Each subject area had different priorities regarding assessment methods. English teachers preferred to use portfolios (48.7%) and teacher questioning and observation (37.2), as contrasted to mathematics teachers who selected students' self-assessments (45.5%) and homework completion (35.7%). A large percentage of science teachers chose pre-and post-tests (46.7%) and teacher-developed assessments (38.1%). Two-thirds of social studies teachers selected group work, and half of them picked behavior/attitude as an assessment practice that should rank within the top five.

Academic Factors. Tables 29-32 show teachers' responses, by subject area, of the academic factors that contributed to end-of-course grades. Teachers across all subject areas indicated that most academic factors were considered when determining grades. Teachers were also consistent in their practice of excluding norm-referenced comparisons.

English teachers indicated that classwork and projects, as well as students' mastery of specific learning objectives had the most influence on grades assigned. Each of these academic factors had the majority of teacher responses in the categories quite a bit or extensive (Table 29).

Table 29

Academic Factors for Determining End-of-Course Grades by Subject Area: English

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	1 (1.6)	5 (7.9)	17 (27.0)	28 (44.4)	12 (19.0)
Daily homework	2 (3.2)	16 (25.4)	26 (41.3)	15 (23.8)	4 (6.3)
Extra credit – academic	12 (19.0)	23 (36.5)	27 (42.9)	1 (1.6)	0 (0)
Improvement of performance	9 (14.3)	8 (12.7)	31 (49.2)	14 (22.2)	1 (1.6)
Inclusion of zeros	12 (19.0)	17 (27.0)	20 (31.7)	14 (22.2)	0 (0)
Mastery of specific learning objectives	1 (1.6)	2 (3.2)	10 (15.9)	28 (44.4)	22 (34.9)
Performance compared with peers	32 (50.8)	13 (20.6)	14 (22.2)	4 (6.3)	0 (0)
Projects	1 (1.6)	8 (12.7)	16 (25.4)	29 (46.0)	9 (14.3)
Quiz scores	1 (1.6)	1 (1.6)	15 (23.8)	37 (58.7)	9 (14.3)
Test scores	1 (1.6)	0 (0)	13 (20.6)	31 (49.2)	18 (28.6)

Table 30 shows the academic factors mathematics teachers report to most influence the determination of grades. Quiz scores and test scores have significant weight when mathematics teachers determine students' grades.

Table 30

Academic Factors for Determining End-of-Course Grades by Subject Area: Mathematics

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	2 (3.9)	0 (0)	24 (47.1)	24 (47.1)	1 (2.0)
Daily homework	0 (0)	9 (17.6)	22 (43.1)	19 (37.3)	1 (2.0)
Extra credit – academic	11 (21.6)	26 (51.0)	13 (25.5)	1 (2.0)	0 (0)
Improvement of performance	8 (16.0)	8 (16.0)	24 (48.0)	10 (20.0)	0 (0)
Inclusion of zeros	9 (17.6)	17 (33.3)	16 (31.4)	7 (13.7)	2 (3.9)
Mastery of specific learning objectives	0 (0)	3 (5.9)	10 (19.6)	25 (49.0)	13 (25.5)
Performance compared with peers	25 (49.0)	14 (27.5)	11 (21.6)	1 (2.0)	0 (0)
Projects	3 (5.9)	7 (13.7)	20 (39.2)	18 (35.3)	3 (5.9)
Quiz scores	0 (0)	1 (2.0)	5 (9.8)	34 (66.7)	11 (21.6)
Test scores	0 (0)	1 (2.0)	4 (7.8)	28 (54.9)	18 (35.3)

Like mathematics teachers, science teachers allowed quiz and test scores to have quite a bit or extensive influence when determining students' grades; however, they also indicated that the student's mastery of specific learning objectives was quite a bit or extensively influential (Table 31).

Table 31

Academic Factors for Determining End-of-Course Grades by Subject Area: Science

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	0 (0)	0 (0)	9 (37.5)	7 (29.2)	8 (33.3)
Daily homework	1 (4.2)	4 (16.7)	11 (45.8)	3 (12.5)	5 (20.8)
Extra credit – academic	4 (16.7)	12 (50.0)	8 (33.3)	0 (0)	0 (0)
Improvement of performance	5 (20.8)	5 (20.8)	9 (37.5)	4 (16.7)	1 (4.2)
Inclusion of zeros	2 (8.7)	4 (17.4)	9 (39.1)	6 (26.1)	2 (8.7)
Mastery of specific learning objectives	1 (4.3)	1 (4.3)	3 (13.0)	12 (52.2)	6 (26.1)
Performance compared with peers	10 (45.5)	9 (40.9)	2 (9.1)	0 (0)	1 (4.5)
Projects	0 (0)	1 (4.2)	10 (41.7)	7 (29.2)	6 (25.0)
Quiz scores	0 (0)	2 (8.3)	4 (16.7)	11 (45.8)	7 (29.2)
Test scores	0 (0)	0 (0)	3 (13.0)	12 (52.2)	8 (34.8)

It is interesting to note in Table 32 that, unlike teachers of other subject areas, all social studies teachers reported using projects to determine students' grades, in addition to those academic factors selected by teachers of other subjects: mastery of specific learning objectives, quiz scores, and test scores. One hundred percent of social studies teachers indicated that they allowed projects to have either *some, quite a bit,* or an *extensive* amount of influence on their end-of-course grades.

Table 32

Academic Factors for Determining End-of-Course Grades by Subject Area: Social Studies

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	0 (0)	1 (4.2)	8 (33.3)	12 (50.0)	3 (12.5)
Daily homework	1 (4.2)	2 (8.3)	12 (50.0)	7 (29.2)	2 (8.3)
Extra credit – academic	4 (16.7)	9 (37.5)	11 (45.8)	0 (0)	0 (0)
Improvement of performance	0 (0)	3 (12.5)	11 (45.8)	8 (33.3)	2 (8.3)
Inclusion of zeros	3 (12.5)	3 (12.5)	10 (41.7)	4 (16.7)	4 (16.7)
Mastery of specific learning objectives	0 (0)	0 (0)	6 (25.0)	12 (50.0)	6 (25.0)
Performance compared with peers	12 (50.0)	5 (20.8)	3 (12.5)	3 (12.5)	1 (4.2)
Projects	0 (0)	0 (0)	8 (33.3)	12 (50.0)	4 (16.7)
Quiz scores	0 (0)	0 (0)	7 (30.4)	12 (52.2)	4 (17.4)
Test scores	0 (0)	0 (0)	5 (20.8)	11 (45.8)	8 (33.3)

Non-Academic Factors.

Non-academic factors are generally student behaviors that impact students' grades – usually in a negative way – but that do not accurately portray the students' academic knowledge or ability. According to measurement experts, non-academic factors should not be included when determining grades (Ebel & Frisbie, 1991; Guskey, 1994). Results of this study show that teachers tended to pick and choose, among the non-academic factors provided, those they wish to consider. Extra credit not related to academics carried no influence at all for the majority of teachers in all subjects, while both student effort and participation were factors included by teachers across the content areas. As reported in Table 33, English teachers indicated that the following non-academic factors carried minimum to no influence on their assigning of grades: attendance in class and behavior/attitude in class.

Table 33

Non-Academic Factors for Determining End-of-Course Grades by Subject Area: English

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	17 (27.0)	19 (30.2)	14 (22.2)	9 (14.3)	4 (6.3)
Behavior/attitude in class	20 (31.7)	16 (25.4)	18 (28.6)	7 (11.1)	2 (3.2)
Extra credit – non-academic	32 (50.8)	23 (36.5)	7 (11.1)	1 (1.6)	0 (0)
Student effort	6 (9.5)	5 (7.9)	28 (44.4)	22 (34.9)	2 (3.2)
Participation	5 (7.9)	9 (14.3)	25 (39.7)	21 (33.3)	3 (4.8)
Work habits (neatness, etc.)	13 (20.6)	20 (31.7)	20 (31.7)	8 (12.7)	2 (3.2)

Table 34 shows that mathematics teachers allowed almost all non-academic factors to carry *some* influence; the exceptions were extra credit and, to a lesser extent, behavior/attitude in class.

Table 34

Non-Academic Factors for Determining End-of-Course Grades by Subject Area: Mathematics

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	17 (33.3)	9 (17.6)	14 (27.5)	8 (15.7)	3 (5.9)
Behavior/attitude in class	20 (39.2)	12 (23.5)	9 (17.6)	8 (15.7)	2 (3.9)
Extra credit – non-academic	29 (58.0)	18 (36.0)	3 (6.0)	0 (0)	0 (0)
Student effort	2 (3.9)	12 (23.5)	25 (49.0)	11 (21.6)	1 (2.0)
Participation	5 (9.8)	14 (27.5)	23 (45.1)	8 (15.7)	1 (2.0)
Work habits (neatness, etc.)	16 (31.4)	15 (29.4)	16 (31.4)	4 (7.8)	0 (0)

Participation and student effort were heavily considered by the majority of science teachers; work habits and attendance in class are also non-academic factors that were considered by about half of the teachers responding. These responses are listed in Table 35 below.

Table 35

Non-Academic Factors for Determining End-of-Course Grades by Subject Area: Science

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	11 (45.8)	3 (12.5)	5 (20.8)	3 (12.5)	2 (8.3)
Behavior/attitude in class	10 (41.7)	8 (33.3)	3 (12.5)	1 (4.2)	2 (8.3)
Extra credit – non-academic	15 (62.5)	7 (29.2)	2 (8.3)	0 (0)	0 (0)
Student effort	4 (16.7)	2 (8.3)	7 (29.2)	9 (37.5)	2 (8.3)
Participation	2 (8.3)	3 (12.5)	11 (45.8)	6 (25.0)	2 (8.3)
Work habits (neatness, etc.)	4 (17.4)	9 (39.1)	5 (21.7)	3 (13.0)	2 (8.7)

Like the results for science teachers, social studies teachers also reported that participation and student effort are large influences on their grading practices. Table 36 shows that just over 83.0% of social studies teachers consider participation (some, quite a bit, and extensive) when assigning grades, while almost 88.0% factor in student effort.

Table 36

Non-Academic Factors for Determining End-of-Course Grades by Subject Area: Social Studies

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	6 (25.0)	5 (20.8)	3 (12.5)	8 (33.3)	2 (8.3)
Behavior/attitude in class	3 (12.5)	10 (41.7)	3 (12.5)	6 (25.0)	2 (8.3)
Extra credit – non-academic	14 (58.3)	6 (25.0)	4 (16.7)	0 (0)	0 (0)
Student effort	1 (4.2)	2 (8.3)	9 (37.5)	9 (37.5)	3 (12.5)
Participation	2 (8.3)	2 (8.3)	12 (50.0)	6 (25.0)	2 (8.3)
Work habits (neatness, etc.)	5 (20.8)	8 (33.3)	7 (29.2)	3 (12.5)	1 (4.2)

Types of Questions. Tables 37-40 show teachers' responses, by subject area, to how often certain types of questions and assessment methods were used. As evidenced below, English and science teachers reported that they "frequently" used essays, matching, multiple-choice and short-answer questions, student exhibits, and portfolios; mathematics and science teachers reported "frequently" using multiple choice and short answer. True-false questions were seldom to never used by the majority of teachers across all subject areas.

Table 37 shows English teachers' responses; the results indicate that English teachers used essays and short-answer questions more frequently than teachers of other subject areas. They also used student exhibits and portfolios more often than others.

Table 37

Frequencies for Types of Questions by Subject Area: English

Types of Questions	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	0 (0)	3 (4.8)	20 (32.3)	31 (50.0)	8 (12.9)
Matching	4 (6.5)	8 (12.9)	24 (38.7)	22 (35.5)	4 (6.5)
Multiple Choice	0 (0)	2 (3.2)	13 (21.0)	39 (62.9)	8 (12.9)
Short Answer	0 (0)	2 (3.2)	13 (21.0)	39 (62.9)	8 (12.9)
True-False	12 (19.4)	24 (38.7)	12 (19.4)	10 (16.1)	4 (6.5)
Student Exhibits	0 (0)	6 (9.7)	17 (27.4)	33 (53.2)	6 (9.7)
Portfolios	4 (6.6)	8 (13.1)	14 (23.0)	30 (49.2)	5 (8.2)

Mathematics teachers used a variety of assessment methods, but tended to use multiple-choice and short-answer questions more regularly than other types of assessments (Table 38).

Table 38

Frequencies for Types of Questions by Subject Area: Mathematics

Types of Questions	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	22 (44.0)	15 (30.0)	11 (22.0)	2 (4.0)	0 (0)
Matching	2 (3.9)	17 (33.3)	24 (47.1)	8 (15.7)	0 (0)
Multiple Choice	0 (0)	6 (11.8)	20 (39.2)	23 (45.1)	2 (3.9)
Short Answer	2 (3.9)	8 (15.7)	12 (23.5)	25 (49.0)	4 (7.8)
True-False	10 (19.6)	24 (47.1)	15 (29.4)	2 (3.9)	0 (0)
Student Exhibits	3 (5.9)	15 (29.4)	23 (45.1)	9 (17.6)	1 (2.0)
Portfolios	11 (22.0)	17 (34.0)	14 (28.0)	7 (14.0)	1 (2.0)

As shown in Table 39, science teachers used a variety of assessment types regularly. Each type listed was used by a large percentage of teachers at least occasionally, with the exception of true-false questions. Multiple-choice questions appeared to be the preferred assessment method, as 69.6% of respondents indicated using them on a weekly basis.

Table 39

Frequencies for Types of Questions by Subject Area: Science

Types of Questions	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	2 (8.7)	4 (17.4)	8 (34.8)	8 (34.8)	1 (4.3)
Matching	0 (0)	2 (9.1)	9 (40.9)	10 (45.5)	1 (4.5)
Multiple Choice	0 (0)	0 (0)	4 (17.4)	16 (69.6)	3 (13.0)
Short Answer	0 (0)	1 (4.3)	7 (30.4)	13 (56.5)	2 (8.7)
True-False	5 (23.8)	5 (23.8)	7 (33.3)	3 (14.3)	1 (4.8)
Student Exhibits	0 (0)	1 (4.5)	10 (45.5)	9 (40.9)	2 (9.1)
Portfolios	2 (8.7)	6 (26.1)	7 (30.4)	7 (30.4)	1 (4.3)

Table 40 shows social studies teachers indicated they were less likely to use portfolios as an assessment method. All social studies teachers responding relied on multiple-choice type questions; all respondents indicated they used them at least weekly.

Table 40

Frequencies for Types of Questions by Subject Area: Social Studies

Types of Questions	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	3 (12.5)	8 (33.3)	4 (16.7)	9 (37.5)	0 (0)
Matching	2 (8.3)	4 (16.7)	8 (33.3)	7 (29.2)	3 (12.5)
Multiple Choice	0 (0)	0 (0)	0 (0)	15 (62.5)	9 (37.5)
Short Answer	1 (4.2)	4 (16.7)	7 (29.2)	10 (41.7)	2 (8.3)
True-False	9 (37.5)	7 (29.2)	5 (20.8)	1 (4.2)	2 (8.3)
Student Exhibits	2 (8.3)	4 (16.7)	8 (33.3)	7 (29.2)	3 (12.5)
Portfolios	8 (33.3)	6 (25.0)	4 (16.7)	2 (8.3)	4 (16.7)

Results by Grade Level. Tables 41-43 show the assessment practices, selected by each grade level, in order from most to least selected. Despite differences in the ranking order, teachers consistently chose among the same options for the assessment methods that should be included in the top five. *Portfolios* was ranked number one by 39.3% of sixth-grade teachers; 33.3% ranked *projects* second; and *homework completion* was chosen as the third-ranked assessment method by 57.1% of respondents. Seventh-grade teachers also ranked *portfolios* as the number one assessment method (41.4%); however, differences in

ranking appeared with the second- (*teacher questioning and observations*) and third- (*grades on assignments*) ranked choices, selected by 33.3% and 37.5% of teachers respectively. Eighth-grade teachers were the only group to rank *teacher developed assessments* (ranked number one by 39.1%) and *class participation* (ranked third by 40.9%) in the top three; the second-ranked assessment method selected by eighth-grade teachers was *teacher questioning and observations*, which was chosen by 26.5% of respondents for the rank of number two.

Table 41

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods by Grade Level: 6th

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
Teacher-developed assessments (*n* = 39)	10 (25.6)	11 (28.2)	8 (20.5)	6 (15.4)	4 (10.3)
Grades on assignments (*n* = 34)	6 (17.6)	6 (17.6)	8 (23.5)	7 (20.6)	5 (14.7)
Portfolios (*n* = 28)	11 (39.3)	7 (25.0)	1 (3.6)	4 (14.3)	3 (10.7)
Teacher questioning and observations (*n* = 26)	4 (15.4)	6 (23.1)	6 (23.1)	5 (19.2)	4 (15.4)
Projects/reports (*n* = 24)	4 (16.7)	8 (33.3)	4 (16.7)	3 (12.5)	3 (12.5)
Pre/post-tests (unit) (*n* = 21)	7 (33.3)	1 (4.8)	4 (19.0)	5 (23.8)	1 (4.8)
End of lesson review (independent) (*n* = 16)	4 (25.0)	1 (6.3)	2 (12.5)	2 (12.5)	4 (25.0)
Self-assessments (students') (*n* = 15)	1 (6.7)	1 (6.7)	7 (46.7)	3 (20.0)	2 (13.3)
District assessments/ benchmarks tests (*n* = 15)	1 (6.7)	3 (20.0)	1 (6.7)	1 (6.7)	6 (40.0)
Class participation (*n* = 12)	0 (0)	2 (16.7)	0 (0)	5 (41.7)	2 (16.7)
State assessments (SOLs) (*n* = 11)	0 (0)	0 (0)	1 (9.1)	3 (27.3)	5 (45.5)
End of lesson review (teacher-led) (*n* = 11)	0 (0)	1 (9.1)	2 (18.2)	1 (9.1)	4 (36.4)
Homework completion /quality (*n* = 7)	0 (0)	0 (0)	4 (57.1)	0 (0)	1 (14.3)
Group work (*n* = 7)	0 (0)	0 (0)	0 (0)	2 (28.6)	2 (28.6)
Behavior/attitude in class (*n* = 4)	0 (0)	1 (25.0)	0 (0)	0 (0)	0 (0)

[b] Percentages calculated based on number of respondents selecting the assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Table 42

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods by Grade Level: 7th

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
Teacher questioning and observations (*n* = 42)	5 (11.9)	14 (33.3)	8 (19.0)	7 (16.7)	8 (19.0)
Teacher-developed assessments (*n* = 40)	14 (35.0)	8 (20.0)	6 (15.0)	4 (10.0)	8 (20.0)
Grades on assignments (*n* = 32)	6 (18.8)	8 (25.0)	12 (37.5)	2 (6.3)	3 (9.4)
Pre/post-tests (unit) (*n* = 30)	11 (36.7)	6 (20.0)	6 (20.0)	3 (10.0)	0 (0)
Portfolios (*n* = 29)	12 (41.4)	3 (10.3)	6 (20.7)	3 (10.3)	1 (3.4)
Projects/reports (*n* = 25)	1 (4.0)	6 (24.0)	2 (8.0)	6 (24.0)	6 (24.0)
District assessments/ benchmarks tests (*n* = 21)	3 (14.3)	1 (4.8)	3 (14.3)	6 (28.6)	4 (19.0)
State assessments (SOLs) (*n* = 21)	0 (0)	3 (14.3)	1 (4.8)	3 (14.3)	9 (42.9)
Class participation (*n* = 17)	3 (17.6)	5 (29.4)	2 (11.8)	2 (11.8)	3 (17.6)
End of lesson review (independent) (*n* = 17)	1 (5.9)	1 (5.9)	2 (11.8)	6 (35.3)	3 (17.6)
Self-assessments (students') (*n* = 16)	0 (0)	0 (0)	2 (12.5)	7 (43.8)	3 (18.8)
Group work (*n* = 13)	0 (0)	0 (0)	1 (7.7)	3 (23.1)	4 (30.8)
End of lesson review (teacher-led) (*n* = 12)	0 (0)	1 (8.3)	4 (33.3)	2 (16.7)	1 (8.3)
Homework completion /quality (*n* = 11)	1 (9.1)	1 (9.1)	1 (9.1)	0 (0)	4 (36.4)
Behavior/attitude in class (*n* = 8)	0 (0)	0 (0)	1 (12.5)	3 (37.5)	0 (0)

[b] Percentages calculated based on number of respondents selecting the assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Table 43

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods by Grade Level: 8th

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
Teacher-developed assessments (*n* = 46)	18 (39.1)	12 (26.1)	7 (15.2)	5 (10.9)	3 (6.5)
Grades on assignments (*n* = 38)	11 (28.9)	9 (23.7)	7 (18.4)	6 (15.8)	3 (7.9)
Teacher questioning and observations (*n* = 34)	5 (14.7)	9 (26.5)	5 (14.7)	9 (26.5)	3 (8.8)
Projects/reports (*n* = 28)	3 (10.7)	5 (17.9)	5 (17.9)	4 (14.3)	7 (25.0)
Pre/post-tests (unit) (*n* = 28)	4 (14.3)	7 (25.0)	3 (10.7)	7 (25.0)	3 (10.7)
Portfolios (*n* = 26)	9 (34.6)	2 (7.7)	5 (19.2)	2 (7.7)	4 (15.4)
Class participation (*n* = 22)	3 (13.6)	0 (0)	9 (40.9)	3 (13.6)	3 (13.6)
State assessments (SOLs) (*n* = 20)	1 (5.0)	3 (15.0)	1 (5.0)	3 (15.0)	5 (25.0)
Homework completion /quality (*n* = 18)	0 (0)	1 (5.6)	4 (22.2)	5 (27.8)	3 (16.7)
End of lesson review (independent) (*n* = 16)	0 (0)	2 (12.5)	2 (12.5)	2 (12.5)	4 (25.0)
Self-assessments (students') (*n* = 16)	0 (0)	2 (12.5)	3 (18.8)	2 (12.5)	3 (18.8)
District assessments/ benchmarks tests (*n* = 15)	0 (0)	1 (6.7)	1 (6.7)	2 (13.3)	4 (26.7)
Behavior/attitude in class (*n* = 13)	0 (0)	1 (7.7)	0 (0)	1 (7.7)	3 (23.1)
End of lesson review (teacher-led) (*n* = 12)	0 (0)	0 (0)	2 (16.7)	2 (16.7)	4 (33.3)
Group work (*n* = 11)	0 (0)	0 (0)	0 (0)	1 (9.1)	2 (18.2)

[b] Percentages calculated based on number of respondents selecting the assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Academic Factors. Factors considered in determining grades were divided into three categories (e.g., academic factors, non-academic factors, and types of questions) and analyzed by grade level (e.g., 6th grade, 7th grade, and 8th grade). Tables 44-46 show that results were similar when responses were analyzed by grade level. All grades had the majority of teachers responding that norm-referenced assessments have *no influence at all* on students' end-of-course grades, and teachers considered a little from each academic factor.

Table 44

Academic Factors for Determining End-of-Course Grades by Grade Level: 6th

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	0 (0)	3 (6.1)	18 (36.7)	22 (44.9)	6 (12.2)
Daily homework	1 (2.0)	7 (14.3)	25 (51.0)	13 (26.5)	3 (6.1)
Extra credit – academic	9 (18.4)	21 (42.9)	19 (38.8)	0 (0)	0 (0)
Improvement of performance	5 (10.4)	7 (14.6)	24 (50.0)	10 (20.8)	2 (4.2)
Inclusion of zeros	8 (16.3)	10 (20.4)	20 (40.8)	8 (16.3)	3 (6.1)
Mastery of specific learning objectives	0 (0)	2 (4.1)	11 (22.4)	24 (49.0)	12 (24.5)
Performance compared with peers	25 (52.1)	10 (20.8)	7 (14.6)	4 (8.3)	2 (4.2)
Projects	0 (0)	6 (12.2)	16 (32.7)	23 (46.9)	4 (8.2)
Quiz scores	0 (0)	2 (4.1)	10 (20.4)	30 (61.2)	7 (14.3)
Test scores	0 (0)	0 (0)	11 (22.9)	24 (50.0)	13 (27.1)

Table 45

Academic Factors for Determining End-of-Course Grades by Grade Level: 7th

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	0 (0)	3 (5.2)	19 (32.8)	25 (43.1)	11 (19.0)
Daily homework	0 (0)	16 (27.6)	21 (36.2)	15 (25.9)	6 (10.3)
Extra credit – academic	9 (15.5)	27 (46.6)	21 (36.2)	1 (1.7)	0 (0)
Improvement of performance	10 (17.2)	9 (15.5)	22 (37.9)	15 (25.9)	2 (3.4)
Inclusion of zeros	14 (24.6)	12 (21.1)	16 (28.1)	13 (22.8)	2 (3.5)
Mastery of specific learning objectives	0 (0)	4 (7.0)	7 (12.3)	28 (49.1)	18 (31.6)
Performance compared with peers	30 (52.6)	10 (17.5)	14 (24.6)	3 (5.3)	0 (0)
Projects	1 (1.7)	6 (10.3)	18 (31.0)	24 (41.4)	9 (15.5)
Quiz scores	0 (0)	1 (1.8)	13 (22.8)	29 (50.9)	14 (24.6)
Test scores	0 (0)	1 (1.7)	9 (15.5)	26 (44.8)	22 (37.9)

Table 46

Academic Factors for Determining End-of-Course Grades by Grade Level: 8th

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	1 (1.8)	2 (3.6)	21 (38.2)	24 (43.6)	7 (12.7)
Daily homework	3 (5.5)	8 (14.5)	24 (43.6)	17 (30.9)	3 (5.5)
Extra credit – academic	14 (25.5)	21 (38.2)	19 (34.5)	1 (1.8)	0 (0)
Improvement of performance	7 (12.7)	8 (14.5)	29 (52.7)	11 (20.0)	0 (0)
Inclusion of zeros	4 (7.3)	18 (32.7)	20 (36.4)	10 (18.2)	3 (5.5)
Mastery of specific learning objectives	2 (3.6)	0 (0)	10 (18.2)	26 (47.3)	17 (30.9)
Performance compared with peers	24 (43.6)	21 (38.2)	9 (16.4)	1 (1.8)	0 (0)
Projects	3 (5.5)	4 (7.3)	19 (34.5)	20 (36.4)	9 (16.4)
Quiz scores	1 (1.8)	1 (1.8)	8 (14.5)	35 (63.6)	10 (18.2)
Test scores	1 (1.8)	0 (0)	5 (9.1)	32 (58.2)	17 (30.9)

Non-Academic Factors. Results indicate that teachers engaged in "hodgepodge" grading, that is, they incorporated a variety of non-academic factors when calculating students' grades so that the one grade is being used to communicate multiple messages.

The majority of teachers across all grades responded that both student effort and participation were considered (e.g., *some, quite a bit*, or *extensively*) when determining grades: student effort (77.6%, 81.0%, and 76.4% for grades six through eight, respectively) and participation (69.4%, 70.7%, and 79.9% respectively). About half of all respondents in grades six, seven, and eight reported that non-acadmeic extra credit carried *no influence at all*. Sixth-grade teachers were more likely to allow students' behaviors/attitudes to carry an *extensive* amount of influence (10.2%) than teachers in grades seven (3.4%) and grade eight (1.8%). Tables 47-49 show the responses, by grade level, teachers gave regarding how influential non-academic factors were when determining end-of-course grades.

Table 47

Non-Academic Factors for Determining End-of-Course Grades by Grade Level: 6th

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	12 (24.5)	12 (24.5)	13 (26.5)	9 (18.4)	3 (6.1)
Behavior/attitude in class	19 (38.8)	11 (22.4)	8 (16.3)	6 (12.2)	5 (10.2)
Extra credit – non-academic	28 (57.1)	17 (34.7)	4 (8.2)	0 (0)	0 (0)
Student effort	4 (8.2)	7 (14.3)	20 (40.8)	16 (32.7)	2 (4.1)
Participation	4 (8.2)	11 (22.4)	20 (40.8)	12 (24.5)	2 (4.1)
Work habits (neatness, etc.)	14 (28.6)	16 (32.7)	13 (26.5)	4 (8.2)	2 (4.1)

Table 48

Non-Academic Factors for Determining End-of-Course Grades by Grade Level: 7th

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	21 (36.2)	16 (27.6)	7 (12.1)	8 (13.8)	6 (10.3)
Behavior/attitude in class	16 (27.6)	21 (36.2)	10 (17.2)	9 (15.5)	2 (3.4)
Extra credit – non-academic	31 (54.4)	20 (35.1)	5 (8.8)	1 (1.8)	0 (0)
Student effort	5 (8.6)	6 (10.3)	22 (37.9)	20 (34.5)	5 (8.6)
Participation	5 (8.6)	12 (20.7)	20 (34.5)	16 (27.6)	5 (8.6)
Work habits (neatness, etc.)	14 (24.6)	21 (36.8)	13 (22.8)	6 (10.5)	3 (5.3)

Table 49

Non-Academic Factors for Determining End-of-Course Grades by Grade Level: 8th

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	18 (32.7)	8 (14.5)	16 (29.1)	11 (20.0)	2 (3.6)
Behavior/attitude in class	19 (34.5)	14 (25.5)	14 (25.5)	7 (12.7)	1 (1.8)
Extra credit – non-academic	32 (58.2)	16 (29.1)	7 (12.7)	0 (0)	0 (0)
Student effort	5 (9.1)	8 (14.5)	27 (49.1)	14 (25.5)	1 (1.8)
Participation	6 (10.9)	5 (9.1)	30 (54.5)	13 (23.6)	1 (1.8)
Work habits (neatness, etc.)	11 (20.0)	15 (27.3)	21 (38.2)	8 (14.5)	0 (0)

Types of Questions. Teachers used a variety of questions/assessment types, based on the responses in Tables 50-52. True-false questions were used the least by all grade levels but especially with sixth-grade teachers. Only two indicated using them more than occasionally, as compared with 12 seventh-grade and nine eighth-grade teachers. Seventh-grade teachers indicated the most frequent use of portfolios, and eighth-grade teachers used essays and student exhibits more often than those at the other grade levels.

Table 50

Frequencies for Types of Questions by Grade Level: 6th

Types of Questions	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	11 (22.9)	6 (12.5)	14 (29.2)	15 (31.3)	2 (4.2)
Matching	4 (8.3)	10 (20.8)	19 (39.6)	13 (27.1)	2 (4.2)
Multiple Choice	0 (0)	2 (4.2)	12 (25.0)	29 (60.4)	5 (10.4)
Short Answer	1 (2.1)	6 (12.5)	12 (25.0)	25 (52.1)	4 (8.3)
True-False	10 (20.8)	26 (54.2)	10 (20.8)	1 (2.1)	1 (2.1)
Student Exhibits	0 (0)	8 (16.7)	19 (39.6)	17 (35.4)	4 (8.3)
Portfolios	4 (8.7)	5 (10.9)	16 (34.8)	16 (34.8)	5 (10.9)

Table 51

Frequencies for Types of Questions by Grade Level: 7th

Types of Questions	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	8 (14.3)	14 (25.0)	14 (25.0)	18 (32.1)	2 (3.6)
Matching	1 (1.8)	10 (17.5)	20 (35.1)	22 (38.6)	4 (7.0)
Multiple Choice	0 (0)	2 (3.5)	11 (19.3)	34 (59.6)	10 (17.5)
Short Answer	2 (3.5)	5 (8.8)	17 (29.8)	28 (49.1)	5 (8.8)
True-False	13 (22.8)	17 (29.8)	15 (26.3)	9 (15.8)	3 (5.3)
Student Exhibits	3 (5.4)	10 (17.9)	18 (32.1)	21 (37.5)	4 (7.1)
Portfolios	9 (15.8)	15 (26.3)	12 (21.1)	17 (29.8)	4 (7.0)

Table 52

Frequencies for Types of Questions by Grade Level: 8th

Types of Questions	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	8 (14.5)	10 (18.2)	16 (29.1)	16 (29.1)	5 (9.1)
Matching	3 (5.6)	11 (20.4)	26 (48.1)	12 (22.2)	2 (3.7)
Multiple Choice	0 (0)	4 (7.3)	14 (25.5)	30 (54.5)	7 (12.7)
Short Answer	0 (0)	4 (7.3)	10 (18.2)	35 (63.6)	6 (10.9)
True-False	13 (24.5)	17 (32.1)	14 (26.4)	6 (11.3)	3 (5.7)
Student Exhibits	2 (3.6)	8 (14.5)	20 (36.4)	21 (38.2)	4 (7.3)
Portfolios	12 (21.8)	17 (30.9)	10 (18.2)	14 (25.5)	2 (3.6)

Results by Ability Level. Factors considered in determining grades were also analyzed by student ability level. Teachers of students in all ability levels selected the following assessment methods at rates that placed them in the top five methods: *teacher-developed assessments*, *grades on assignments*, and *portfolios*. There were differences between ability levels regarding the remaining assessment methods selected by the greater number of respondents: teachers of students in comprehensive classes, as well as those of honors/gifted classes, selected *teacher questioning and observations*; comprehensive

teachers and inclusive/collaborative teachers agreed that *pre-post tests* should be among the top five; but inclusive/collaborative teachers were the only group to select *district assessments/benchmark tests* at a rate that placed them in the top five. Ranking choices also differed by ability levels. Percentages represent those who selected it for that specific ranking. Teachers of comprehensive students ranked portfolios first (41.5%), projects/reports second (36.7%), homework completion/quality third (50.0%), self-assessments by students fourth (30.0%), and state assessments/SOLs fifth (51.9%);. Honors/gifted teachers selected pre/post-tests first (35.5%), teacher questioning and observations second (30.8%), teacher-led end of lesson review third and fifth (28.6% each), and project/reports and homework completion/quality tied for fourth (21.1%). Teachers of inclusive/collaborative students ranked portfolios first (70.0%), teacher-developed assessments second (33.3%), pre/post-tests third (50.0%), group work fourth (66.7%), and district assessments/benchmark tests fifth (66.7%). Tables 53-55 show assessment methods rankings for each ability level.

Table 53

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods by Ability Level: Comprehensive

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
Teacher-developed assessments (*n* = 61)	24 (39.3)	15 (24.6)	9 (14.8)	8 (13.1)	5 (8.2)
Grades on assignments (*n* = 58)	12 (20.7)	14 (24.1)	15 (25.9)	10 (17.2)	6 (10.3)
Teacher questioning and observations (*n* = 56)	6 (10.7)	16 (28.6)	9 (16.1)	13 (23.2)	10 (17.9)
Portfolios (*n* = 41)	17 (41.5)	6 (14.6)	6 (14.6)	4 (9.8)	3 (7.3)
Pre/post-tests (unit) (*n* = 36)	9 (25.0)	4 (11.1)	6 (16.7)	9 (25.0)	4 (11.1)
Projects/reports (*n* = 30)	3 (10.0)	11 (36.7)	4 (13.3)	4 (13.3)	4 (13.3)
District assessments/ benchmarks tests (*n* = 29)	2 (6.9)	4 (13.8)	3 (10.3)	6 (20.7)	9 (31.0)
State assessments (SOLs) (*n* = 27)	0 (0)	3 (11.1)	2 (7.4)	4 (14.8)	14 (51.9)
Class participation (*n* = 25)	2 (8.0)	4 (16.0)	6 (24.0)	6 (24.0)	4 (16.0)
Self-assessments (students') (*n* = 20)	0 (0)	0 (0)	7 (35.0)	6 (30.0)	3 (15.0)
End of lesson review (independent) (*n* = 20)	5 (25.0)	0 (0)	3 (15.0)	2 (10.0)	5 (25.0)
End of lesson review (teacher-led) (*n* = 18)	0 (0)	2 (11.1)	3 (16.7)	3 (16.7)	5 (27.8)
Homework completion /quality (*n* = 14)	1 (7.1)	0 (0)	7 (50.0)	1 (7.1)	3 (21.4)
Behavior/attitude in class (*n* = 11)	0 (0)	2 (18.2)	0 (0)	3 (27.3)	2 (18.2)
Group work (*n* = 10)	0 (0)	0 (0)	1 (10.0)	1 (10.0)	3 (30.0)

[b] Percentages calculated based on number of respondents selecting the assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Table 54

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods by Ability Level: Honors/Gifted

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
Teacher-developed assessments (*n* = 52)	17 (32.7)	12 (23.1)	10 (19.2)	7 (13.5)	5 (9.6)
Teacher questioning and observations (*n* = 39)	6 (15.4)	12 (30.8)	8 (20.5)	8 (20.5)	3 (7.7)
Projects/reports (*n* = 38)	4 (10.5)	6 (15.8)	6 (15.8)	8 (21.1)	9 (23.7)
Grades on assignments (*n* = 37)	9 (24.3)	7 (18.9)	9 (24.3)	4 (10.8)	4 (10.8)
Portfolios (*n* = 31)	7 (22.6)	5 (16.1)	6 (19.4)	3 (9.7)	5 (16.1)
Pre/post-tests (unit) (*n* = 31)	11 (35.5)	7 (22.6)	1 (3.2)	5 (16.1)	0 (0)
Class participation (*n* = 24)	4 (16.7)	3 (12.5)	5 (20.8)	3 (12.5)	4 (16.7)
End of lesson review (independent) (*n* = 24)	0 (0)	3 (12.5)	3 (12.5)	5 (20.8)	6 (25.0)
State assessments (SOLs) (*n* = 22)	1 (4.5)	2 (9.1)	1 (4.5)	4 (18.2)	5 (22.7)
Self-assessments (students') (*n* = 20)	1 (5.0)	3 (15.0)	4 (20.0)	3 (15.0)	3 (15.0)
District assessments/ benchmarks tests (*n* = 19)	2 (10.5)	1 (5.3)	2 (10.5)	3 (15.8)	3 (15.8)
Homework completion /quality (*n* = 19)	0 (0)	1 (5.3)	2 (10.5)	4 (21.1)	4 (21.1)
Group work (*n* = 18)	0 (0)	0 (0)	0 (0)	3 (16.7)	5 (27.8)
End of lesson review (teacher-led) (*n* = 14)	0 (0)	0 (0)	4 (28.6)	1 (7.1)	4 (28.6)
Behavior/attitude in class (*n* = 13)	0 (0)	0 (0)	1 (7.7)	1 (7.7)	1 (7.7)

[b] Percentages calculated based on number of respondents selecting the assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Table 55

Frequency Distribution of Respondents' Ranking of the Top Five Assessment Methods by Ability Level: Inclusive/Collaborative

Assessment Methods[c]	Ranking #1 *n* (%)	Ranking #2 *n* (%)	Ranking #3 *n* (%)	Ranking #4 *n* (%)	Ranking #5 *n* (%)[b]
District assessments/ benchmarks tests (*n* = 14)	0 (0)	0 (0)	0 (0)	0 (0)	2 (66.7)
Teacher-developed assessments (*n* = 12)	2 (16.7)	4 (33.3)	1 (8.3)	0 (0)	5 (41.7)
Pre/post-tests (unit) (*n* = 12)	2 (16.7)	3 (25.0)	6 (50.0)	1 (8.3)	0 (0)
Portfolios (*n* = 10)	7 (70.0)	1 (10.0)	0 (0)	2 (20.0)	0 (0)
Grades on assignments (*n* = 10)	2 (20.0)	2 (20.0)	4 (40.0)	1 (10.0)	1 (10.0)
Projects/reports (*n* = 9)	1 (11.1)	1 (11.1)	1 (11.1)	2 (22.2)	3 (33.3)
Teacher questioning and observations (*n* = 7)	2 (28.6)	2 (28.6)	2 (28.6)	0 (0)	1 (14.3)
Self-assessments (students') (*n* = 7)	0 (0)	0 (0)	1 (14.3)	3 (42.9)	2 (28.6)
End of lesson review (independent) (*n* = 4)	0 (0)	1 (25.0)	0 (0)	2 (50.0)	0 (0)
Group work (*n* = 3)	0 (0)	0 (0)	0 (0)	2 (66.7)	0 (0)
State assessments (SOLs) (*n* = 3)	0 (0)	1 (33.3)	0 (0)	1 (33.3)	0 (0)
Homework completion /quality (*n* = 3)	0 (0)	1 (33.3)	0 (0)	0 (0)	1 (33.3)
End of lesson review (teacher-led) (*n* = 3)	0 (0)	0 (0)	1 (33.3)	1 (33.3)	0 (0)
Class participation (*n* = 3)	0 (0)	0 (0)	0 (0)	1 (33.3)	1 (33.3)
Behavior/attitude in class (*n* = 1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

[b] Percentages calculated based on number of respondents selecting the assessment method.
[c]Totals may not add up across rows because some respondents ranked all, rather than the top five.

Academic Factors. A variety of academic factors have influence on students' end-of-course grades, regardless of ability level (see Tables 56-58). Mastery of specific learning objectives, quiz scores, and test scores carry the most influence for the majority of teachers.

Non-Academic Factors. Tables 59-61 present the responses of teachers, based on students' ability levels, regarding the influence of non-academic factors on end-of-course grades. Teachers of students at the comprehensive and honors/gifted levels reported similar practices; both groups of teachers considered several non-academic factors and allowed those non-academic factors to carry *some influence*. Teachers of students in inclusive/collaborative classes tended to respond more in the *no influence at all* category for such non-academic factors as attendance in class, behavior/attitude in class, and extra credit.

Table 56

Academic Factors for Determining End-of-Course Grades by Ability Level: Comprehensive

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	1 (1.2)	3 (3.6)	24 (28.9)	38 (45.8)	17 (20.5)
Daily homework	1 (1.2)	16 (19.3)	36 (43.4)	21 (25.3)	9 (10.8)
Extra credit – academic	14 (16.9)	38 (45.8)	30 (36.1)	1 (1.2)	0 (0)
Improvement of performance	10 (12.2)	13 (15.9)	38 (46.3)	20 (24.4)	1 (1.2)
Inclusion of zeros	18 (21.7)	19 (22.9)	29 (34.9)	12 (14.5)	5 (6.0)
Mastery of specific learning objectives	1 (1.2)	3 (3.6)	14 (16.9)	39 (47.0)	26 (31.3)
Performance compared with peers	40 (48.8)	21 (25.6)	16 (19.5)	5 (6.1)	0 (0)
Projects	3 (3.6)	10 (12.0)	30 (36.1)	31 (37.3)	9 (10.8)
Quiz scores	1 (1.2)	1 (1.2)	14 (17.1)	47 (57.3)	19 (23.2)
Test scores	1 (1.2)	0 (0)	12 (14.5)	42 (50.6)	28 (33.7)

Table 57

Academic Factors for Determining End-of-Course Grades by Ability Level: Honors/Gifted

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	0 (0)	4 (6.3)	28 (44.4)	25 (39.7)	6 (9.5)
Daily homework	2 (3.2)	13 (20.6)	25 (39.7)	20 (31.7)	3 (4.8)
Extra credit – academic	15 (23.8)	26 (41.3)	21 (33.3)	1 (1.6)	0 (0)
Improvement of performance	8 (12.7)	8 (12.7)	32 (50.8)	12 (19.0)	3 (4.8)
Inclusion of zeros	6 (9.7)	15 (24.2)	25 (40.3)	13 (21.0)	3 (4.8)
Mastery of specific learning objectives	0 (0)	2 (3.2)	12 (19.4)	30 (48.4)	18 (29.0)
Performance compared with peers	27 (43.5)	17 (27.4)	13 (21.0)	3 (4.8)	2 (3.2)
Projects	1 (1.6)	4 (6.3)	17 (27.0)	29 (46.0)	12 (19.0)
Quiz scores	0 (0)	2 (3.2)	13 (20.6)	36 (57.1)	12 (19.0)
Test scores	0 (0)	1 (1.6)	8 (12.9)	31 (50.0)	22 (35.5)

Table 58

Academic Factors for Determining End-of-Course Grades by Ability Level: Inclusive/Collaborative

Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Classwork	0 (0)	1 (6.3)	5 (31.3)	9 (56.3)	1 (6.3)
Daily homework	1 (6.3)	2 (12.5)	9 (56.3)	4 (25.0)	0 (0.0)
Extra credit – academic	3 (18.8)	6 (37.5)	7 (43.8)	0 (0)	0 (0)
Improvement of performance	5 (31.3)	3 (18.8)	5 (31.3)	3 (18.8)	0 (0)
Inclusion of zeros	2 (12.5)	7 (43.8)	2 (12.5)	5 (31.3)	0 (0)
Mastery of specific learning objectives	1 (6.3)	1 (6.3)	2 (12.5)	9 (56.3)	3 (18.8)
Performance compared with peers	12 (75.0)	3 (18.8)	1 (6.3)	0 (0)	0 (0)
Projects	0 (0)	2 (12.5)	7 (43.8)	6 (37.5)	1 (6.3)
Quiz scores	0 (0)	1 (6.3)	3 (18.8)	12 (75.0)	0 (0)
Test scores	0 (0)	0 (0)	4 (25.0)	10 (62.5)	2 (12.5)

Table 59

Non-Academic Factors for Determining End-of-Course Grades by Ability Level: Comprehensive

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	23 (27.7)	20 (24.1)	16 (19.3)	18 (21.7)	6 (7.2)
Behavior/attitude in class	29 (34.9)	19 (22.9)	17 (20.5)	13 (15.7)	5 (6.0)
Extra credit – non-academic	54 (65.1)	21 (25.3)	8 (9.6)	0 (0)	0 (0)
Student effort	7 (8.4)	11 (13.3)	33 (39.8)	28 (33.7)	4 (4.8)
Participation	8 (9.6)	11 (13.3)	36 (43.4)	22 (26.5)	6 (7.2)
Work habits (neatness, etc.)	27 (32.5)	17 (20.5)	24 (28.9)	11 (13.3)	4 (4.8)

Table 60

Non-Academic Factors for Determining End-of-Course Grades by Ability Level: Honors/Gifted

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	19 (30.2)	13 (20.6)	18 (28.6)	8 (12.7)	5 (7.9)
Behavior/attitude in class	17 (27.0)	24 (38.1)	12 (19.0)	7 (11.1)	3 (4.8)
Extra credit – non-academic	27 (42.9)	27 (42.9)	8 (12.7)	1 (1.6)	0 (0)
Student effort	4 (6.3)	7 (11.1)	31 (49.2)	17 (27.0)	4 (6.3)
Participation	4 (6.3)	13 (20.6)	28 (44.4)	16 (25.4)	2 (3.2)
Work habits (neatness, etc.)	8 (12.9)	28 (45.2)	18 (29.0)	7 (11.3)	1 (1.6)

Table 61

Non-Academic Factors for Determining End-of-Course Grades by Ability Level: Inclusive/Collaborative

Non-Academic Factors	No Influence At All *n* (%)	Minimum Influence *n* (%)	Some Influence *n* (%)	Quite a Bit of Influence *n* (%)	Extensive Influence *n* (%)
Attendance in class	10 (62.5)	3 (18.8)	1 (6.3)	2 (12.5)	0 (0)
Behavior/attitude in class	8 (50.0)	2 (12.5)	4 (25.0)	2 (12.5)	0 (0)
Extra credit – non-academic	9 (60.0)	6 (40.0)	0 (0)	0 (0)	0 (0)
Student effort	3 (18.8)	3 (18.8)	5 (31.3)	5 (31.3)	0 (0)
Participation	3 (18.8)	4 (25.0)	6 (37.5)	3 (18.8)	0 (0)
Work habits (neatness, etc.)	4 (25.0)	7 (43.8)	5 (31.3)	0 (0)	0 (0)

Types of Questions. As shown in Tables 62-64, the frequency of the use of various types of questions was similar among teachers of all ability levels; they used a variety but trended less towards true-false and more towards multiple choice. When looking at the three assessment types that result in the creation of a product (essays, student exhibits, and portfolios), and when focusing on the responses for *occasionally* and *frequently*, teachers of comprehensive students used essays (63.4%), student exhibits (75.6%), and portfolios (57.6%) more often than teachers of honors/gifted (52.5%, 71.0%, and 48.4%) and inclusive/collaborative students (56.3%, 66.6%, and 56.3%).

Table 62

Frequencies for Types of Questions by Ability Level: Comprehensive

Types of Questions	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	14 (17.1)	14 (17.1)	30 (36.6)	22 (26.8)	2 (2.4)
Matching	6 (7.3)	16 (19.5)	31 (37.8)	24 (29.3)	5 (6.1)
Multiple Choice	0 (0)	3 (3.7)	16 (19.5)	52 (63.4)	11 (13.4)
Short Answer	2 (2.4)	8 (9.8)	19 (23.2)	46 (56.1)	7 (8.5)
True-False	22 (27.2)	31 (38.3)	16 (19.8)	7 (8.6)	5 (6.2)
Student Exhibits (reports and projects)	3 (3.7)	14 (17.1)	33 (40.2)	29 (35.4)	3 (3.7)
Portfolios	9 (11.3)	18 (22.5)	25 (31.3)	21 (26.3)	7 (8.8)

Table 63

Frequencies for Types of Questions by Ability Level: Honors/Gifted

Types of Questions	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	11 (18.0)	11 (18.0)	10 (16.4)	22 (36.1)	7 (11.5)
Matching	2 (3.2)	11 (17.7)	29 (46.8)	17 (27.4)	3 (4.8)
Multiple Choice	0 (0)	4 (6.5)	16 (25.8)	32 (51.6)	10 (16.1)
Short Answer	0 (0)	6 (9.7)	14 (22.6)	35 (56.5)	7 (11.3)
True-False	13 (21.3)	22 (36.1)	17 (27.9)	7 (11.5)	2 (3.3)
Student Exhibits (reports and projects)	2 (3.2)	8 (12.9)	20 (32.3)	24 (38.7)	8 (12.9)
Portfolios	12 (19.4)	17 (27.4)	10 (16.1)	20 (32.3)	3 (4.8)

Table 64

Frequencies for Types of Questions by Ability Level: Inclusive/Collaborative

Types of Questions	Never *n* (%)	Seldom (quarterly) *n* (%)	Occasionally (monthly) *n* (%)	Frequently (weekly) *n* (%)	Always (daily) *n* (%)
Essays	2 (12.5)	5 (31.3)	4 (25.0)	5 (31.3)	0 (0)
Matching	0 (0)	4 (26.7)	6 (40.0)	5 (33.3)	0 (0)
Multiple Choice	0 (0)	1 (6.3)	5 (31.3)	9 (56.3)	1 (6.3)
Short Answer	1 (6.3)	1 (6.3)	6 (37.5)	6 (37.5)	2 (12.5)
True-False	1 (6.3)	8 (50.0)	5 (31.3)	2 (12.5)	0 (0)
Student Exhibits (reports and projects)	0 (0)	4 (26.7)	5 (33.3)	5 (33.3)	1 (6.7)
Portfolios	4 (25.0)	2 (12.5)	3 (18.8)	6 (37.5)	1 (6.3)

Relationships Between Assessment and Grading Practices and Student Achievement

The third research question asked if there was a relationship between assessment and grading practices used by middle school, core academic subject teachers and their students' SOL scores and end-of-course grades. Correlations were calculated to determine whether a relationship existed, and, if so, how strong the relationship was, between teachers' assessment and grading practices and students' end-of-course grades as well as

SOL scores. Although it is a common mistake to assume that where correlational relationships exist, these relationships provide evidence of cause and effect, one must be careful not to make such assumptions. Correlations do provide information about the strength of a relationship – if one exists – but they do not indicate the phenomenon of cause and effect (Huck, 2000). Correlations are considered to be low/weak if they are close to zero; low correlations indicate that the variables are independent (Huck, 2000). Correlations are high/strong if they are close to one; correlations that fall at the 0.5 level, positive or negative, are considered to be moderate in nature (Huck, 2000).

Class end-of-course averages were based on a four-point scale, with four representing the letter grade "A"; the closer the class average is to four, the higher the end-of-course average. Averages for the SOL scores were based on the point scale established for the assessments: averages below 300 indicate *below proficiency*; scores from 400 to 499 indicate *pass/proficient*; scores from 500 to 599 indicate *pass advanced*; and a score of 600 is a perfect score. Table 65 shows the end-of-course averages and SOL averages of respondents overall and by subject area. As shown in Table 65, the average end-of-course grade for all subject areas is in the grade-level range of "C," and all SOL averages are in the *pass/proficient* range.

Table 65

Mean Values for Class Average and SOL Average by Subject Area

Subject	Class Average			SOL Average		
	n	$\overline{X}$	*SD*	*n*	$\overline{X}$	*SD*
Overall	139	2.76	.66	128	482.52	43.63
English	55	2.77	.61	54	480.26	38.00
Mathematics	40	2.85	.67	40	489.25	44.61
Science	21	2.66	.72	11	475.27	50.48
Social Studies	22	2.73	.74	22	482.32	51.43

Overall Results

As shown in Table 66, statistically significant, low correlations were evident between end-of-course grades and projects (r = .194), non-academic extra credit (r = .195), and student exhibits (r = .210); there was also a statistically significant, low correlation between SOL averages and projects (r = .209). There was a high correlation between the end-of-course average and SOL average (r = .767), which indicates that when the end-of-course average is high, the SOL average for that course also will be high.

Table 66

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods

Grading Factors and Assessment Methods	End-of-course Average (n = 139)	SOL Average (n = 128)
Academic Factors		
Classwork	-.136	-.125
Daily homework	.035	.013
Extra credit – academic	-.106	-.048
Improvement of performance	.059	-.064
Inclusion of zeros	.098	.144
Mastery of specific learning objectives	-.024	-.011
Performance compared with peers	.000	-.055
Projects	.194*	.209*
Quiz scores	-.065	-.038
Test scores	-.013	.009
Non-Academic Factors		
Attendance in class	-.002	-.075
Behavior/attitude in class	.001	-.048
Extra credit – non-academic	.195*	.072
Student effort	.004	.004
Participation	-.088	-.121
Work habits (neatness, etc.)	.023	.004
Assessment Methods		
Essays	.044	.036
Matching	-.075	-.031
Multiple Choice	-.014	-.033
Short Answer	-.059	-.022
True-False	-.012	.025
Student Exhibits	.210*	.149
Portfolios	.085	-.021
SOL Average	.767**	1.000

*p<0.05 **p<0.01

Results by Subject Area

Tables 67-70 provide the correlation results by subject area. All subjects had a statistically significant correlation between the end-of-course average and the SOL average ($p < .01$). The science correlation was the strongest ($r = .891$), followed by English ($r = .727$), mathematics ($r = .778$), and social studies ($r = .746$).

As shown in Table 67, English results yielded the greatest number of statistically significant correlations with both end-of-course averages and SOL averages. Statistically significant, albeit low, correlations were found between end-of-course average and classwork ($r = .269$), non-academic extra credit ($r = .364$), and student exhibits ($r = .284$); low correlations were also evident between SOL averages and inclusion of zeros ($r = .365$), projects ($r = .344$), essays ($r = .298$), and student exhibits ($r = .419$). Each of these statistically significant relationships indicates that end-of-course averages and SOL averages tended to be higher the more often those factors and/or assessment methods were used, as reported by teachers.

Table 67

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods by Subject Area: English

Grading Factors and Assessment Methods	End-of-course Average (n = 55)	SOL Average (n = 54)
Academic Factors		
Classwork	-.269*	-.176
Daily homework	-.026	.002
Extra credit – academic	-.195	-.108
Improvement of performance	.046	-.091
Inclusion of zeros	.087	.365**
Mastery of specific learning objectives	-.042	.119
Performance compared with peers	.034	.018
Projects	.206	.344*
Quiz scores	-.129	-.047
Test scores	.020	.152
Non-Academic Factors		
Attendance in class	-.003	-.057
Behavior/attitude in class	.087	-.036
Extra credit – non-academic	.364**	.225
Student effort	.064	-.060
Participation	-.096	-.092
Work habits (neatness, etc.)	-.026	-.032
Assessment Methods		
Essays	.196	.298*
Matching	.018	-.083
Multiple Choice	-.056	-.095
Short Answer	.033	.117
True-False	.120	.092
Student Exhibits	.284*	.419**
Portfolios	-.082	.096
SOL Average	.727**	1.000

*p<0.05 **p<0.01

Table 68 shows that in the area of mathematics there were no statistically significant correlations between grading factors and assessment methods and end-of-course averages or SOL averages.

Table 68

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods by Subject Area: Mathematics

Grading Factors and Assessment Methods	End-of-course Average (n = 40)	SOL Average (n = 40)
Academic Factors		
Classwork	-.034	-.077
Daily homework	.196	.186
Extra credit – academic	-.014	-.003
Improvement of performance	-.071	-.165
Inclusion of zeros	.033	.064
Mastery of specific learning objectives	-.044	-.149
Performance compared with peers	-.258	-.120
Projects	.195	.048
Quiz scores	.080	-.096
Test scores	-.052	-.180
Non-Academic Factors		
Attendance in class	-.109	-.152
Behavior/attitude in class	-.124	-.074
Extra credit – non-academic	-.072	-.119
Student effort	-.055	.049
Participation	-.076	-.147
Work habits (neatness, etc.)	-.051	-.043
Assessment Methods		
Essays	-.197	-.204
Matching	-.162	-.047
Multiple Choice	.000	-.063
Short Answer	-.101	-.127
True-False	-.095	.026
Student Exhibits	.150	.056
Portfolios	-.016	-.300
SOL Average	.778**	1.000

*p<0.05 **p<0.01

Science yielded two additional correlations, all moderate and with end-of-course averages: inclusion of zeros (r = .454) and student exhibits (r = .549). Complete results are reported in Table 69. It is important to note that the science SOL is only administered in the 8th grade; therefore, the sample size is very small (n = 11), and results may not generalize to the broader teaching population at the district level.

Table 69

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods by Subject Area: Science

Grading Factors and Assessment Methods	End-of-course Average (n = 21)	SOL Average (n = 11)
Academic Factors		
Classwork	-.219	-.051
Daily homework	-.199	-.054
Extra credit – academic	-.056	-.088
Improvement of performance	.341	.250
Inclusion of zeros	.454*	.261
Mastery of specific learning objectives	.042	.267
Performance compared with peers	.234	-.086
Projects	.176	.501
Quiz scores	-.102	.464
Test scores	-.159	.338
Non-Academic Factors		
Attendance in class	.008	-.423
Behavior/attitude in class	-.251	-.375
Extra credit – non-academic	.122	-.113
Student effort	-.002	.049
Participation	.148	.107
Work habits (neatness, etc.)	.004	.112
Assessment Methods		
Essays	.154	.190
Matching	-.261	-.133
Multiple Choice	-.099	.252
Short Answer	-.227	-.107
True-False	.041	.191
Student Exhibits	.549**	.436
Portfolios	.364	.347
SOL Average	.891**	1.000

*p<0.05 **p<0.01

Social studies results indicated one additional correlation, at the moderate level, with end-of-course average: projects (r = .471). See Table 70.

Table 70

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods by Subject Area Social Studies

Grading Factors and Assessment Methods	End-of-course Average (n = 22)	SOL Average (n = 22)
Academic Factors		
Classwork	.243	-.033
Daily homework	.244	-.136
Extra credit – academic	-.175	-.059
Improvement of performance	-.023	-.122
Inclusion of zeros	.027	-.038
Mastery of specific learning objectives	-.041	-.270
Performance compared with peers	.088	-.116
Projects	.471*	.301
Quiz scores	-.162	-.360
Test scores	.055	-.205
Non-Academic Factors		
Attendance in class	.125	.045
Behavior/attitude in class	.266	-.013
Extra credit – non-academic	.265	.109
Student effort	-.030	-.035
Participation	-.294	-.275
Work habits (neatness, etc.)	.307	.102
Assessment Methods		
Essays	.418	.420
Matching	.039	.161
Multiple Choice	.385	.249
Short Answer	.052	.129
True-False	-.191	-.111
Student Exhibits	.221	.066
Portfolios	.382	.149
SOL Average	.746**	1.000

*p<0.05 **p<0.01

Results by Grade Level

Table 71 details the statistical, descriptive information on class and SOL averages specific to grade levels and including overall data. As shown in Table 71, the class average for all grade levels was in the "C" range and all SOL averages were in the 400s, which is considered to be a *pass/proficient* level of performance.

Table 71

Mean Values for Class Average and SOL Average by Grade Level

	Class Average			SOL Average		
	n^e	$\overline{X}$	*SD*	n^e	$\overline{X}$	*SD*
Overall	139	2.76	.66	128	482.52	43.63
6th grade	39	2.82	.59	35	472.03	43.91
7th grade	51	2.72	.65	45	487.49	41.34
8th grade	48	2.77	.73	47	486.04	45.44

[e] Totals may not equal the overall *n* because some respondents did not identify the grade level.

When analyzed by grade level, statistically significant correlations were found at all grades between the end-of-course average and the SOL average ($p < .01$): 6th grade ($r = .793$), 7th grade ($r = .779$), and 8th grade ($r = .795$).

Table 72 shows additional statically significant correlations in sixth-grade. Low correlations were found between projects and the end-of-course average ($r = .387$, $p < .05$) as well as between SOL averages ($p < .05$) and projects ($r = .369$), participation ($r = -.368$), and matching ($r = -.338$).

Table 72

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods by Grade Level: 6th

Grading Factors and Assessment Methods	End-of-course Average (n = 39)	SOL Average (n = 35)
Academic Factors		
Classwork	-.145	-.101
Daily homework	-.054	-.070
Extra credit – academic	-.209	-.176
Improvement of performance	.073	-.133
Inclusion of zeros	-.064	.059
Mastery of specific learning objectives	-.074	-.029
Performance compared with peers	.081	-.083
Projects	.387*	.369*
Quiz scores	-.123	-.283
Test scores	-.145	-.181
Non-Academic Factors		
Attendance in class	.157	.040
Behavior/attitude in class	-.029	.021
Extra credit – non-academic	.027	-.049
Student effort	.023	-.068
Participation	-.233	-.368*
Work habits (neatness, etc.)	.069	.004
Assessment Methods		
Essays	-.077	-.140
Matching	-.159	-.338*
Multiple Choice	-.238	-.159
Short Answer	-.172	-.045
True-False	-.165	-.070
Student Exhibits	.162	.225
Portfolios	.072	.063
SOL Average	.793**	1.000

*p<0.05 **p<0.01

As shown in Table 73, there was one additional statistically significant correlation at the seventh-grade level: a low, negative correlation between end-of-course average and classwork (r = -.283, p < .05).

Table 73

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods by Grade Level: 7th

Grading Factors and Assessment Methods	End-of-course Average (n = 51)	SOL Average (n = 45)
Academic Factors		
Classwork	-.283*	-.219
Daily homework	-.050	-.104
Extra credit – academic	-.259	-.095
Improvement of performance	.066	.011
Inclusion of zeros	.106	.090
Mastery of specific learning objectives	-.025	-.130
Performance compared with peers	-.020	-.023
Projects	.008	.023
Quiz scores	-.138	-.106
Test scores	-.062	-.032
Non-Academic Factors		
Attendance in class	-.018	-.107
Behavior/attitude in class	.081	-.032
Extra credit – non-academic	.256	-.070
Student effort	.146	.215
Participation	.021	-.044
Work habits (neatness, etc.)	-.113	-.184
Assessment Methods		
Essays	.083	.041
Matching	-.046	.005
Multiple Choice	-.051	-.158
Short Answer	-.028	-.034
True-False	.015	-.022
Student Exhibits	.166	.037
Portfolios	-.020	-.166
SOL Average	.779**	1.000

*p<0.05 **p<0.01

Correlations calculated for the eighth-grade revealed one additional statistically significant relationship; there was a low correlation between the end-of-course average ($p < .05$) and student exhibits ($r = .290$). All correlations are found in Table 74.

Table 74

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods by Grade Level: 8th

Grading Factors and Assessment Methods	End-of-course Average ($n = 48$)	SOL Average ($n = 47$)
Academic Factors		
Classwork	.025	-.081
Daily homework	.162	.157
Extra credit – academic	.097	.064
Improvement of performance	.030	-.080
Inclusion of zeros	.200	.243
Mastery of specific learning objectives	.012	.079
Performance compared with peers	-.076	-.061
Projects	.264	.272
Quiz scores	.045	.155
Test scores	.120	.148
Non-Academic Factors		
Attendance in class	-.108	-.117
Behavior/attitude in class	-.060	-.095
Extra credit – non-academic	.251	.260
Student effort	-.126	-.138
Participation	-.099	-.051
Work habits (neatness, etc.)	.156	.141
Assessment Methods		
Essays	.100	.161
Matching	-.016	.137
Multiple Choice	.168	.126
Short Answer	-.018	.035
True-False	.060	.052
Student Exhibits	.290*	.216
Portfolios	.176	.130
SOL Average	.795**	1.000

*$p<0.05$ **$p<0.01$

Results by Ability Level

Table 75 details the statistical, descriptive information on class and SOL averages specific to ability levels and including overall data. As shown in Table 75, the class average for honors/gifted students was in the "B" range, while the average for comprehensive and inclusive/collaborative was in the "C" range; the SOL average for the honors/gifted students was also higher, in the 500s, which puts it in the *pass advanced* category, but the others were in the 400s, a *pass/proficient* level of performance.

Table 75

Mean Values for Class Average and SOL Average by Ability Level

Ability Level	Class Average			SOL Average		
	n[e]	$\overline{X}$	*SD*	n[e]	$\overline{X}$	*SD*
Overall	139	2.76	.66	128	482.52	43.63
Comprehensive	69	2.36	.55	63	451.06	32.34
Honors/Gifted	59	3.32	.29	54	521.81	19.42
Inclusive/Collaborative	10	2.25	.46	10	461.30	12.72

[e] Totals may not equal the overall *n* because some respondents did not identify the ability level.

When analyzed specific to ability level, a statistically significant correlation was found between the end-of-course average and the SOL average for the comprehensive level only ($r = .624$, $p < .01$). Additional statistically significant correlations also were found at the comprehensive level. As listed in Table 76, there was a low, negative correlation with norm-referencing and both end-of-course average ($r = -.258$, $p < .05$) and SOL average ($r = -.304$, $p < .05$).

Table 76

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods by Ability Level: Comprehensive

Grading Factors and Assessment Methods	End-of-course Average (n = 69)	SOL Average (n = 63)
Academic Factors		
Classwork	-.193	-.011
Daily homework	.072	.046
Extra credit – academic	-.022	.027
Improvement of performance	-.046	-.170
Inclusion of zeros	-.029	-.009
Mastery of specific learning objectives	-.194	-.037
Performance compared with peers	-.258*	-.304*
Projects	-.010	-.047
Quiz scores	-.066	-.035
Test scores	-.119	-.013
Non-Academic Factors		
Attendance in class	.053	-.006
Behavior/attitude in class	.005	-.105
Extra credit – non-academic	-.060	-.154
Student effort	-.085	-.170
Participation	-.063	-.210
Work habits (neatness, etc.)	-.087	-.178
Assessment Methods		
Essays	-.142	-.123
Matching	-.172	-.154
Multiple Choice	.115	.025
Short Answer	-.221	-.161
True-False	-.065	-.042
Student Exhibits	.019	-.106
Portfolios	.167	.206
SOL Average	.624**	1.000

*p<0.05 **p<0.01

Table 77 shows the correlations calculated for classes at the honors/gifted ability level. Statistically significant relationships were found between the end-of-course average and improvement of performance ($r = .393$, $p < .01$) and non-academic extra credit ($r = .287$, $p < .05$).

As shown in Table 78, several statistically significant correlations were found within the inclusive/collaborative ability level. The data in Table 78 is based on a small sample, which should be taken into consideration when examining the results; however, the results have been included for consistency and to highlight any trends.

The inclusion of zeros had a moderate, positive correlation with SOL average ($r = .634$, $p < .05$). End-of- course average correlated with two academic factors, two non-academic factors, and one question/assessment method: classwork ($r = .655$, $p < .05$), improvement of performance ($r = -.797$, $p < .01$), attendance in class ($r = -.849$, $p < .01$), participation ($r = -.671$, $p < .05$), and true-false ($r = -.697$, $p < .05$). It is important to note that classwork was the only positive correlation, an indicator of a positive relationship with end-of-course averages - the more classwork is incorporated into grades, the higher the class average.

Table 77

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods by Ability Level: Honors/Gifted

Grading Factors and Assessment Methods	End-of-course Average (n = 59)	SOL Average (n = 54)
Academic Factors		
Classwork	.228	-.016
Daily homework	.184	.123
Extra credit – academic	-.033	.051
Improvement of performance	.393**	-.163
Inclusion of zeros	.027	.091
Mastery of specific learning objectives	.132	-.142
Performance compared with peers	.154	.036
Projects	.203	-.007
Quiz scores	.022	-.137
Test scores	.090	-.140
Non-Academic Factors		
Attendance in class	.050	-.245
Behavior/attitude in class	.128	-.071
Extra credit – non-academic	.287*	.071
Student effort	.165	-.002
Participation	-.010	.016
Work habits (neatness, etc.)	.061	.060
Assessment Methods		
Essays	.098	-.052
Matching	.100	.024
Multiple Choice	.059	.090
Short Answer	-.018	-.135
True-False	.211	.108
Student Exhibits	.118	-.084
Portfolios	.217	-.211
SOL Average	.067	1.000

*p<0.05 **p<0.01

Table 78

Correlations for End-of-Course and SOL Averages with Grading Factors and Assessment Methods by Ability Level: Inclusive/Collaborative

Grading Factors and Assessment Methods	End-of-course Average (n = 10)	SOL Average (n = 10)
Academic Factors		
Classwork	.655*	.215
Daily homework	.579	.394
Extra credit – academic	.034	.410
Improvement of performance	-.797**	-.066
Inclusion of zeros	-.284	.634*
Mastery of specific learning objectives	-.009	.602
Performance compared with peers	-.285	.257
Projects	-.365	.525
Quiz scores	-.137	.477
Test scores	.038	.506
Non-Academic Factors		
Attendance in class	-.849**	-.250
Behavior/attitude in class	-.078	-.089
Extra credit – non-academic	.359	-.039
Student effort	-.404	.165
Participation	-.671*	.045
Work habits (neatness, etc.)	.571	-.035
Assessment Methods		
Essays	.357	.491
Matching	-.257	.066
Multiple Choice	.114	-.337
Short Answer	-.073	-.127
True-False	-.697*	.056
Student Exhibits	-.081	.406
Portfolios	.285	.464
SOL Average	.052	1.000

*p<0.05 **p<0.01

Summary

Purpose of grading. Regardless of how the data were analyzed, teachers agreed on the fundamental purposes of grading. Analyses showed consistent overall perceptions among subject areas, grade levels, and ability levels. When asked to rank the purpose of grades, respondents consistently selected the same primary purposes (ranked one and two): *measurement of student's mastery of academic content* and *provide feedback to students*. Although the same secondary purposes of grading (*communicate to parents, provide incentive to student/motivation*, and *measurement of student's level of effort*) were generally selected by participants, the rank order varied across subjects, grade levels and ability levels. Science teachers, who ranked *measurement of student's level of responsibility* as the fifth purpose of grading, were the only group who did not include the purpose *communicate to parents* in the top five.

Results by grade level and ability level also yielded clearly defined primary purposes of grading. The number one purpose chosen by all grade levels and two of the three ability levels (comprehensive and honors/gifted) was *measurement of student's mastery of academic content*. All also selected the same purposes two and three: *provide feedback to students* and *communicate to parents* respectively. The teachers of students in inclusive/collaborative classes varied from other teachers in their ranking choices and order. The number one purpose of grading for teachers in inclusive/collaborative classrooms was *provide feedback to students*, followed by *measurement of student's mastery of academic content*; both were also selections made by others, but the order was different. Inclusive/collaborative teachers' order implies that communication with students

regarding their progress is the most important function of grades. Teachers of inclusive/collaborative students were also the only group to choose *select, identify, or group students* as one of the primary purposes of grading; they ranked it third, and it was selected in the top five by a large number of inclusive/collaborative teachers.

Attitudes toward grading. To determine teachers' attitudes towards grading, teachers were asked to respond with their level of agreement to several statements about grading practices. Those statements were divided into two categories, non-academic factors and academic factors, according to the grading and assessment literature. The stance of measurement experts is that grades can – and should – only communicate one message: the student's academic progress (Ebel & Frisbie, 1991; Guskey, 1994; Lambating & Allen, 2002). If behavior, attitude, and other factors that are not directly related to mastery of content are to be communicated, then they should be done so using a separate report (Wendel & Anderson, 1994).

Results from teachers' responses showed that teachers do not agree completely with the stance of measurement experts regarding what grades should reflect about the student. The majority of teachers (95.1%) agreed or strongly agreed that grades should reflect students' mastery of the content, which is consistent with measurement experts' recommendations. However, teachers also indicated that grades should reflect progress/improvement over time (92.0%) and students' efforts (81.2%). In addition, measurement experts clearly delineate between formative and summative assessments (Black & Wiliam, 1998b; Brookhart, 2007; Wiliam & Black, 1996), strongly suggesting that only summative assessments should be used for assigning grades (Black & Wiliam,

1998b; Brookhart, 2008; Guskey, 1994). Yet, teachers' responses show a disagreement in their thinking regarding the use of assessments when assigning grades: 77.9% agreed/strongly agreed that homework completion should be included in final grades; 65.0% agreed/strongly agreed that all assignments should count towards the course grade. The results were consistent when analyzed by subject area, grade level, and ability level.

Assessment and grading methods. Similar to the questions that asked respondents to rank the purposes of grading for research question one, teachers were also asked to rank certain assessment methods, choosing only what they believed to be the primary methods for determining students' grades. Consistently found in the top three assessment methods were *teacher-developed assessments, assignment of grades,* and *teacher questioning and observations*.

For research question number two, teachers were also asked to indicate the amount of influence various factors had on the end-of-course grades they assigned to students. Factors were grouped into two categories: academic factors or non-academic factors. Results indicated that teachers used a variety of academic factors and question/assessment types to determine end-of-course grades. Results of analyses conducted by subject area, grade level, and ability level indicated that teachers used several academic factors at various levels of influence. Respondents also indicated using several non-academic factors when determining end-of-course grades, and there was widespread agreement in the non-academic factors they chose to use: *student effort* and *participation* but not *non-academic extra credit*. Teachers were split in their use of the non-academic factors *attendance in class* and *behavior/attitude in class* when arriving at grades.

Relationship to student achievement. Statistically significant, positive correlations were found between end-of-course averages and SOL averages for all subject areas and all grade levels, which indicates that a higher end-of-course average is associated with a higher SOL average. When calculated across all subject areas, grade levels, and ability levels, statistically significant, positive correlations were found between end-of-course averages and projects, non-academic extra credit, and student exhibits. Under this scenario, there was only one statistically significant positive relationship found involving the SOL average, and that was with the use of projects.

Data were disaggregated by the subject area, and unlike the other subject areas that had an additional one or two statistically significant correlations, several statistically significant relationships, positively and negatively correlated, were found for English teachers. Positive correlations were detected between end-of-course average and non-academic extra credit as well as the use of student exhibits; a negative correlation was found with incorporating classwork into the grade. Also in English, positive correlations were found between SOL averages and inclusion of zeros, projects, essays, and student exhibits when determining grades. Aside from the statistically significant relationship between the end-of-course average and the SOL average, mathematics was the only subject area not to have had any additional statistically significant correlations between grading and assessment practices and the student achievement measures.

Several statistically significant relationships were found among grade-level correlational analyses that were specific to grade levels but not true for all grades, with the exception being the strong, positively correlated, statistically significant relationship

between end-of-course averages and SOL averages. Sixth grade was the only grade level to result in statistically significant relationships with projects (to end-of-course average and SOL average), participation (negative correlation with SOL average), and matching (negative correlation with SOL average). Seventh grade had one additional statistically significant correlation: a negative relationship between classwork and end-of-course average. There also was only one additional statistically significant relationship in eighth grade, found between student exhibits and end-of-course averages.

For the sake of consistency and to highlight any trends that may have appeared, results were also disaggregated by ability level, despite the low number of participants for the inclusive/collaborative category (n = 10). Results from correlations calculated by ability level differed from those based on subject area and grade level. Unlike the others, not all ability levels yielded a statistically significant relationship between end-of-course average and SOL average; such a relationship was found only at the comprehensive ability level. This may have been due to sample size and a lack of variability in the student achievement measures, especially for the honors/gifted group. In addition, there were very few – and some surprising – statistically significant relationships found with the ability level correlations. Comprehensive classes yielded negatively correlated, statistically significant relationships between norm-referenced criteria and both end-of-course average and SOL average; honors/gifted classes had a positively correlated, statistically significant relationship between improvement of performance and end-of-course average as well as between non-academic extra credit and end-of-course average; inclusive/collaborative classes had the greatest number of statistically significant relationships present: with end-

of-course average, there was a positive correlation with classwork, but negative correlations with improvement of performance, attendance in class, participation, and use of true-false questions. There was also a positive correlation between inclusion of zeros and SOL average.

Analyses of the data show, overall, that teachers in this district generally agreed on the primary purposes of grading. They also used a variety of assessment methods and were consistent with the non-academic factors they chose to consider when determining students' grades. The results of this study also found statistically significant relationships for all subject areas between students' end-of-course averages and SOL averages and relationship with the use of certain grading and assessment practices with student achievement.

Chapter 5

Conclusion and Recommendations

Overview

The purpose of this study was to examine teachers' grading and assessment practices in an effort to answer the question "What do grades mean?" A descriptive, non-experimental, quantitative study that used an electronic questionnaire to survey middle school, core academic subject teachers in a large, suburban school district in Virginia was developed and administered. The research questions that guided the study design and procedures were as follows:

1. What do middle school, core academic subject teachers report to be the purpose of grades?
2. What assessment and grading practices do middle school, core academic subject teachers use most often in their classrooms when determining students' grades?
3. Is there a relationship between assessment and grading practices used by middle school, core academic subject teachers and students' SOL scores and end-of-course grades?

A secondary focus of this study was to examine the extent to which middle school, core academic subject teachers' self-reported assessment methods and grading practices differed by subject area, grade level, and student ability level.

To examine teachers' survey responses, frequencies were used to determine what teachers considered to be the primary purposes of grading, teachers' attitudes toward grading, and teachers' assessment methods and factors considered with determining grades. Correlations were also used to determine if statistically significant relationships existed between students' end-of-course averages and/or SOL averages and various assessment and grading practices reported by teachers.

Since the administration of this survey, the school district has made changes to several of its policies, including its policy on grading and teachers' responsibilities for grading. The large, suburban school district in which the survey was administered recently changed its school board policy on grading, so that it is now a *written* policy that grades should reflect the student's mastery of academic objectives. The survey for this study was administered prior to the change in board policy, so consequently the new policy would not have influenced teachers' responses to the survey items.

Significant Findings

Purpose of grading. Teachers generally agreed on the primary purposes of grading. The majority of teachers responding believe that the number one purpose of grading is to serve as a tool that measures students' mastery of academic content. They also agreed that the second purpose of grades is to communicate to students about their progress. When it came to delineating the secondary purposes of grading (rankings three,

four, and five), the results were not as clear-cut regarding rank-order; however, teachers were consistent in the selections made while ranking the top five purposes: measurement of student's mastery of academic content, provide feedback to students, communicate to parents, provide incentive to student/motivation, and measurement of student's level of effort.

These results were consistent across subject area, grade level, and student ability level, with one exception. Teachers of students in inclusive/collaborative classes, those classes comprised of both students of average academic ability and students receiving services under special education, responded differently. Inclusive/collaborative teachers were the only group to have a large enough number of teachers choosing *select, identify, or group students* so that it was listed within the top five most selected purposes, and it ranked third among purposes with one-third of inclusive/collaborative teachers so ranking it. This was the start of a trend in the data; if there was an exception to the choices indicated by respondents, it would be with teachers of students in inclusive/collaborative classes.

Attitudes toward grading. An analysis of the responses showed that teachers across subject areas, grade levels, and ability levels share similar attitudes towards grading: grades should reflect the student's mastery of the content (95.1%), but they should also reflect student's progress/improvement (92.0%) and effort (81.2%) and should include grades from formative activities such as homework completion (77.9%). The difference in how teachers responded is found in the results from teachers of students in inclusive/collaborative classes: the percentage of teachers who agreed/strongly agreed that

grades should reflect student's mastery of the content was over 90.0% for all, except those who teach in an inclusive/collaborative setting (87.5%); two-thirds of inclusive/collaborative teachers agreed that grades should reflect progress/improvement, while greater than 85.0% of teachers in all other disaggregated categories agreed/strongly agreed with that statement. These responses support results from previous studies that have found teachers' practices are not aligned with recommendations found in research literature on appropriate grading practices regarding what information should be included and what message should be communicated by grades. This disparity results in a question regarding the applicability of research recommendations to real-life practice.

Assessment and grading practices. Respondents were asked to rank their top five choices out of 15 listed for their preferred assessment methods; the choices provided had been used previously in a study by Guskey (2007) in which he asked participants to rank all fifteen. Focusing on the number of respondents selecting the assessment method as a top-five preference, results showed that teachers preferred using the following as indicators of student achievement: teacher-developed assessments, grades on assignments, teacher questioning and observation, portfolios, and pre/post-tests. The response rate for the use of portfolios was surprising; it ranked first as an assessment method, but it was selected by teachers at a rate that placed it fourth on the list. The researcher expected projects/reports to rank higher, but based on both percentages for rankings and the number of teachers choosing it and its position on the list, it was not included in the top five choices. Results were similar when the data were disaggregated by subject area; however, projects/reports

did make the top five list, replacing pre/post-tests for English teachers and portfolios for social studies teachers.

This section of the survey also asked teachers to indicate the level of influence various factors have on students' end-of-course grades. Items for this portion of the survey consisted of factors that had been included in and adapted from previous studies (Cauley et. al, 2008; Frary et. al, 1993); the scale of influence was adapted from the Cauley et. al (2008) study. For the analysis of the results, the factors were separated into three categories: academic factors, non-academic factors, and types of questions. The grouping of items into the categories was based on the recommendations made by measurement experts as reported in the literature on acceptable grading and assessment practices.

Results showed that teachers across all subject areas, grade levels, and student ability levels allow a variety of academic factors to have varied levels of influence. It is interesting to note that almost half of all teachers, regardless of how the data were disaggregated, indicated that norm-referenced criteria (e.g., comparing students to their peers) had no influence at all when they determined end-of-course grades. Teachers also indicated that they allowed test scores and quiz scores to influence heavily the end-of-course grades they assign to students. Despite agreeing, almost unanimously, that grades should reflect student's mastery of academic content, it was surprising to find that mastery of specific learning objectives did not carry more influence (29.0% of all teachers selected *extensive influence*). Subject-area results were 35.9% English, 25.5% mathematics, 26.1% science, and 25.0% social studies. Grade-level results were 24.5% 6^{th} grade, 31.6% 7^{th} grade, and 30.9% 8^{th} grade. Ability-level results were 31.3% comprehensive, 29.0%

honors/gifted, and 18.8% inclusive/collaborative. This disparity between belief and practice reinforces findings by other researchers that teachers' practices in the classroom are not aligned with the best practices recommended by measurement specialists, even when they know what research says is a best practice (Brookhart, 1993; Cross & Frary, 1999) and, as is the case with this study, even when they believe that certain factors should be included (Baron, 2000). For example, teachers generally agreed that mastery of academic content is the message that grades should include; however, they also reported that mastery of academic content was given some consideration, as opposed to quite a bit or an extensive amount of influence, when determining grades.

Relationship to student achievement. Correlations were calculated to examine the relationships between the various grading factors and question/assessment types and the end-of-course averages and SOL averages. In addition to analyzing responses by subject area and grade level, data also were disaggregated by student ability level and reported in the results of this study for consistency and to highlight trends that may have occurred; however, the sample size for correlations based on student ability level was small due to SOL data not being available for all classes.

Those items that measurement experts assert should be excluded from consideration when assigning students' grades can be used by teachers to get a good understanding of how students are progressing; however, non-academic factors (e.g., behavior, motivation, participation, etc.) should not be incorporated into the grade assigned. The correlations resulted in only one statistically significant relationship between a non-academic factor and SOL average: participation ($r = -.368$, $p < .05$).

Findings show that there is a statistically significant relationship, positively correlated at a high level between the end-of-course average and the SOL average; this is true for data that were analyzed prior to being disaggregated, as well as for data organized by subject area and by grade level. Results were surprising in that there were not more statistically significant relationships found, as well as in some of the statistically significant relationships that were found. It was surprising to find that data from English teachers had a larger number of statistically significant relationships than any other group (e.g., a whole-group analysis, an analysis by subject area, grade level, and student ability level); previous studies found that mathematics teachers were more likely to grade strictly on academic progress (Pilcher, 1994; Terwilliger, 1966) and using non-academic factors less that teachers of other subjects (McMillan, 2001), while English teachers were more like to grade based on their beliefs about students' abilities (Pilcher, 1994).

It is also interesting to note that with correlations calculated on types of questions/assessment methods, there were some statistically significant relationships found between those that required students to create items (e. g., essays, student exhibits, and portfolios) and end-of-course averages and/or SOL averages. Brookhart (2007) noted that teachers were hesitant to move away from assessments that require students to apply their prior knowledge to new situations, preferring instead to assess students using materials formatted similarly to standardized tests. The relationships found in this study with types of questions/assessment methods and student achievement measures indicate that teachers should incorporate more activities that require students to use what they have learned and apply that knowledge to new situations.

Discussion

What do middle school, core academic subject teachers report to be the purpose of grades? It is interesting and reassuring to find that teachers overwhelmingly agreed that the primary purpose of grades is to communicate students' levels of mastery of the content objective to students and to parents, based on their choices for the top three primary purposes of grades. However, it is disconcerting to also find that despite their belief that mastery is what should be communicated by grades, their responses also suggest that mastery does not have an extensive influence on the grades that are issued to students.

Results from this study are similar to previous studies that found a discrepancy between teachers' beliefs about what grades mean and the factors teachers consider and include when assigning grades (Baron, 2000; Brookhart, 1993; Cross & Frary, 1999). Teachers participating in this study agreed that the purpose of grades is to communicate a message about the student's mastery of the content; however, they also agree that non-academic factors (e.g., behavior, effort, etc.) should be reported, so they include them when determining end-of-course grades. Similar results were found in studies conducted by Brookhart (1993, 1994), and McMillan (1999).

What assessment and grading practices do middle school, core academic subject teachers use most often in their classrooms when determining students' grades? The issue of "hodgepodge grading" (i.e. using one grade to communicate more than the student's mastery of the content) has been raised numerous times in the literature on grading and assessment practices. Researchers in the areas of grading and assessment, along with measurement experts, hold that there are certain practices teachers should not

engage in when determining grades (e.g., using norm-referenced criteria, giving zeros as grades, grading activities that are formative in nature, and including non-academic factors). Teachers participating in this study indicate they do not use norm-referencing with their students, but they do engage in/agree with certain other practices that defy measurement experts' recommendations: zeros should be included; homework completion and accuracy should be included; and progress/improvement and student effort should be included, as well as using grades to motivate students.

Teachers participating in this study indicated that they use a variety of question types and a variety of assessment methods when measuring students' mastery of the content, which aligns with recommendations of measurement experts (Carlson, 2003; Guskey, 2000). When discussing assessments, the role of district and state assessments invariably enters the conversation. Airasian (1988), Black and Wiliam (1998b), and Stiggins (2002) assert that high stakes tests are not necessarily the best indicators of student achievement. Teachers participating in this study did not rank district assessments or state assessments as a primary assessment method. This is surprising in light of the fact that the school district requires core academic subject teachers to administer county-created benchmark assessments each quarter as a tool to determine students' mastery of the objectives for that grading period, and that the state requires SOLs to be administered at the end of each academic course. One explanation for teachers' not giving these assessments high priority in their rankings is district policy on how they can use students' scores. Benchmark assessments have been described by the district as a formative assessment tool for teachers to use to determine what content needs to be re-taught. As

such, benchmark scores are not to be incorporated into students' grades. By contrast, the district allows SOL scores for high-school credit courses to be substituted for exam grades only if the score will help the student's end-of-course grade; since there are no exams for middle school courses, SOL scores are not used to supplement students' end-of-course grades. Students' scores on SOL assessments cannot be converted for use as an end-of-course grade; however, some teachers continue to use benchmarks as a summative assessment and incorporate students' scores on the benchmark into the quarterly grade.

Teachers in this study were divided in their views on giving students a zero for work that was not completed. Measurement specialists suggest that zeros should not be assigned because they do not accurately reflect a student's knowledge about a subject (Guskey, 2000, 2002; McMillan, 1999); however, about half of teachers across all subject areas, except English, agreed or strongly agreed that students should receive zeros for incomplete work. In addition, the inclusion of zeros had a statistically significant, positive correlation with SOL average ($p < .01$) for teachers of English and science ($p < .05$) as well as teachers of students in an inclusive/collaborative setting ($p < .05$). This division in thinking regarding the inclusion of zeros may be reflective of a move some schools in the district have made towards implementing some version of a "No Zero Policy," in which teachers require students to make up and/or re-do work until it meets some set criteria. Of those teachers who indicated that their school has a specific grading policy for them to follow, approximately 45.0% stated that their school had some version of a "No Zero Policy." Within schools with such a policy, there are teachers who are still resistant to it because they believe that a "No Zero Policy" is setting students up for future failure and/or

creating a culture of procrastination among students who may take the attitude "I'll do it later." There are also teachers in those same schools who stand firmly behind the "No Zero Policy" as a means of moving towards standardized grading built on the premise of not knowing what the student knows if he/she does not complete the assignment. The variation in school-specific policies regarding assigning zeros may explain the findings of this study.

There are several common arguments as to why teachers feel that all assignments must be graded and why they feel grades should speak to more than mastery of objectives. Teachers in the school where this researcher works believe, and often use the argument, that students will not do the work, or take it seriously, if everything is not graded, so how is one to determine what students really know and can do. This sentiment of using grades as a motivator for students to do the work has been echoed by teachers in other buildings in that same school district and by teachers in other neighboring districts. It also has been discussed in previous studies (Barnes, 1985; Ebel & Frisbie, 1991; Guskey, 2000). This thinking is also a possible explanation for teachers' engaging in grading practices that do not align with teachers' stated beliefs about the primary purpose of grading, and it reinforces the question: "What needs to be done to remove the disconnect between theory and practice?" Brookhart (2007) suggested that teachers' focusing on grading rather than providing feedback may have resulted in the unintended consequence of students working only for grades; findings by Cauley et. al (2008) support this stance. To redirect the focus to the learning that is required, rather than on the grade that is given, teachers need to focus

on the message they intend grades to send and align their grading practices so that the two are connected.

Is there a relationship between assessment and grading practices used by middle school, core academic subject teachers and students' SOL scores and end-of-course grades? This study found that there are statistically significant relationships that are highly correlated between students' end-of-course averages and their SOL averages. That is not to suggest that students have high SOL averages because their class averages are high, but it does indicate that students with high class averages are more likely to have higher averages on the SOL assessments. The SOLs are the standards that students must master in each core content area, which ultimately results in Virginia's schools being held to a standards-based framework for student achievement. This study found that there were few statistically significant relationships found with either academic or non-academic factors and SOL averages, and there were fewer statistically significant relationships found with SOL averages than there were with end-of-course averages. These findings support arguments that such assessments are the ultimate tools for assessing students' achievements (Airasian, 1988; Black and Wiliam, 1998b; Stiggins, 2002).

Greatly surprising were the statistically significant, positive correlations between inclusion of zeros – a practice measurement experts adamantly oppose – and end-of-course average by science teachers ($r = .454$, $p < .05$) and SOL average by English teachers ($r = .365$, $p < .05$). An analysis of the factors English and science teachers indicated having at least "some" influence (e.g., some, quite a bit, or extensive) on the end-of-course grade assigned showed several similarities in their practices. However, striking differences also

appeared, and those differences may be indicators of the statistically significant correlation in English with SOL average and that in science with end-of-course average.

A little more than two-thirds of English teachers who agreed/strongly agreed that students should receive zeros for incomplete work also allowed the inclusion of zeros to carry at least some influence on the end-of-course grade assigned. However, a large percentage of these English teachers were also influenced (e.g., some, quite a bit, or extensive) by students' mastery of specific learning objectives (95.5%) and quiz scores (95.5%). The SOL assessments are standards-based, and if quizzes are developed directly from the standards, then this may explain why a statistically significant relationship was found between the inclusion of zeros and SOL averages. Approximately four out of five science teachers who agreed/strongly agreed that students should receive zeros also indicated that zeros had some influence on the end-of-course grades; however, large percentages of these science teachers also allowed several other factors to influence those grades: 100.0% are influenced by classwork, 83.3% by daily homework and academic extra credit, and 75.0% by student effort and participation.

Further analysis, specific to those who agreed/strongly agreed that zeros should be included (n = 72) and those whose schools have a specific grading policy teachers are required to follow (n = 29), found that 44.8% of those teachers' schools have a "No Zero Policy" in place. Such policy requires teachers to allow students to re-do and/or make-up work and/or that provides the lowest grade a teacher can give a student without actually assigning a zero.

Based on the findings of this study, grading factors that generally require students to produce some artifact (e.g., essays, projects, reports) are more likely to result in higher levels of student achievement. Teachers' responses indicated that they are included in the rotation of assessment methods used, but generally they are not used more than monthly. English teachers were the exception, with the majority stating that they use essays, student exhibits, and portfolios at least weekly. This supports a statement by Brookhart (2007) that speaks to the fear teachers have of moving away from assessment methods that do not challenge students to use their prior knowledge to solve new problems, so they stick to assessment methods that mimic standardized, state and/or district assessments and that are easy to grade.

Recommendations

Implications for practice. The fact that statistically significant relationships were found in all subject areas between end-of-course averages and SOL scores implies that careful consideration should be given to what goes into those end-of-course grades, especially in light of the reality that SOL averages are also used by Virginia to calculate a school's status under the federal No Child Left Behind (NCLB) Act. In light of the heightened emphasis on student achievement, using standardized assessments such as the SOLs, grades are one way to communicate student achievement. It is imperative that the disparity between theory and practice be analyzed and revisited so that best practices match classroom practices for assessments methods and considerations for grading.

This study's finding that teachers believe the primary purpose of grades to be a measurement of student's mastery of academic content, coupled with the re-written school

board policy that mandates students' grades be reflective of students' mastery of the content's objectives, could lead to an examination of the format used by middle schools in this school district to report students' grades. If one also considers that teachers' responses to the survey items indicate that grades reflect more than students' mastery of the standards, but the current reporting systems allows the reporting of one grade, then there is a clear argument for a report card that reflects the standards-based emphasis of the school system and that incorporates spaces for teachers to report on other behaviors that are not standards-based but that teachers believe also impact students' achievement (e. g., attendance, attitude/behavior, effort, etc.). Measurement specialists support reporting on students' progress with non-academic factors as long as that reporting is separate from the grade, which is to be used to communicate only on academic progress (Canady & Hotchkiss, 1989; Guskey, 2002; Terwilliger, 1966; Wendel & Anderson, 1994).

Professional development opportunities for teachers that intimately involve teachers in a reflection on their beliefs about grading and an analysis of the factors they consider when grading could be the beginning of helping teachers to see and understand the disconnect between classroom practices and research recommendations for grades as reflections of students' achievements. Results from this study indicated some differences within groups and between groups, so designing such professional development to target teachers by subject areas, grade-levels, and ability levels would be beneficial for highlighting similarities and differences in current practice and to start the conversation around standardizing practices for fairness to students. Bringing in an educator who has researched and published numerous articles on grading, and who is respected by peers, to

talk with teachers about how certain grading practices are detrimental to student achievement would be an effective start to having teachers re-examine what they do and why they do it when it comes to grading, without the personal and philosophical debates that arise when the focus is on a specific policy, such as the "No Zero Policy."

Future studies. More research is needed in the area of the impact of zeros on students' performance. This study found two instances where there was a statistically significant, positive correlation between the inclusion of zeros and some other factor, specifically end-of-course averages and SOL averages. It would be interesting to learn under what conditions, if any, the assignment of zeros would be acceptable in light of the relationship found in this study and in light of the strong stance among researchers that the issuing of zeros invalidates grades and is detrimental to students' morales.

The discussion on the role of formative assessment and summative assessments is an area that needs further study. Measurement experts disagree as to whether the two can be used together, or if combining the two results in the unintended consequence of students' losing interest in the learning and only focusing on the grade. While conducting this study and analyzing teachers' responses, this researcher often asked if teachers understand the differences between formative and summative assessments. This study did not explain or distinguish between formative and summative assessments; however, the response choices for grading factors considered by teachers when assigning end-of-course grades could fit into either category, depending on how teachers used them. A future study that was designed to measure teachers' understanding of formative versus summative assessments and/or a study that labeled response choices as such would be beneficial to the

literature on assessment practices as well as grading practices. Are teachers cognizant of the types of assessments they are using with students and using to determine grades? Is one used more than the other? Are they intermingled? What is the impact of the type of assessment used on students' academic achievements?

In light of the fact that researchers are usually not practitioners, it is understandable, to a degree, why best practices recommended by measurement experts are not implemented in the classrooms by practicing teachers. However, this study also found that despite their agreement that *measurement of student's mastery of academic content* is the primary purpose of grades, teachers also reported that students' mastery of academic content is not a major influencing factor when they are determining grades. This raises the question, "Why teachers' grading practices do not mirror what they state they believe regarding the purpose of grading?" An area in need of future research is an analysis of what causes the disparity between research and practice and suggestions for bridging that gap. One way to approach such research would be through a comparison of teachers' reported practices with results from observations of their actual practice in the classroom.

Limitations

This study was limited to the middle school, core academic content teachers in one school division in Virginia, which limited the initial sample size. Add in the factors of employee attrition and the fact that the district sponsored several studies at the beginning of the 2009-2010 school year that teachers were asked to participate in prior to the administration of the survey for this research study, and the participation rate was greatly reduced from the limited initial sample size.

Time constraints and the teachers' desire to think back to the previous school year were also areas of limitation for this study. The researcher hoped to have the survey sent out in early October, when teachers are willing to think back to what they did and why in regard to grading and assessment practices during the previous year but also far enough into the new school year that they are relaxed enough with the start-up of a new year to take the time to complete the survey thoughtfully and reflectively. However, the school district was delayed in granting the official permission for the survey to be distributed. This delay, in addition to the number of surveys already received by teachers, may have resulted in teachers choosing not to respond to yet another survey, and one that asked them to think back to the previous school year.

Respondents were asked to select and reference one section of the course they taught during the previous school year throughout the survey, which increased the chances of respondents misreporting information about the section of the class they selected (e. g., student ability level and class period). This research was highly dependent on teachers' self-reporting the information requested, without there being any way to verify that the information provided was accurate. Although the request to reference the same section of the course when responding to questions was repeated throughout the survey, the possibility remained that respondents may have combined practices from various sections. Also, the number of respondents who used a class with the student ability level of *inclusive/collaborative* was such that findings may have been skewed by the small sample size.

Due to the way some schools code their classes in their master schedules, it was not possible for the school system's data specialist to compile end-of-course averages for all classes, resulting in smaller than anticipated samples for data using end-of-course averages. The sample size of science teachers with SOL data was also small because the science SOL is only administered in the eighth grade, so correlations within the subject area of science may have been skewed due to the limited sample.

Conclusions

This non-experimental, quantitative, descriptive study provided information on what teachers believe the primary purposes of grading to be, their attitudes about grading, their assessment methods and grading practices, and the relationship of those methods and practices to student achievement. The study found that teachers are consistent in what they believe the number one purpose of grading to be: a measurement of students' mastery of the academic content. The study also found that teachers believe grades should reflect more than students' content mastery and that teachers use a variety of assessment methods and incorporate various grading practices when determining students' end-of-course grades.

Results from this study are similar to results from previous studies (Brookhart, 1994; Cauley et. al, 2008; Cross & Frary, 1999; Guskey, 2007) that show that teachers include several different factors when assigning students' grades, including factors that measurement experts adamantly avow should not be considered or included in the grade assignment. It would be beneficial to students, parents, and school staff if the results of this study led to the creation of a grade-reporting document that accurately reflected

students' mastery of academic content while also providing teachers with the opportunity to report on other factors that impact students' academic achievement (e.g., attendance, behavior, effort, participation, etc.).

As a result of data collected from this study, statistically significant relationships were found between end-of-course averages and SOL averages as a result of correlational calculations conducted on data specific to each core subject area. The finding of these statistically significant relationships, in all subject areas, addresses a gap in the literature on grading and assessment practices and their impact on student achievement. Perhaps these findings will lead to further studies on which grading practices, and therefore end-of-course averages – have the greatest positive impact on students' achievement, which is currently quantified via SOL averages.

List of References

List of References

Airasian, P. W. (1988). The case of state-mandated, high-stakes testing. *Educational Evaluation and Policy Analysis, 10*(4), 301-313. Retrieved February 7, 2009, from http://www.jstor.org/stable/1164172

Airasian, P. W. & Jones, A. M. (1993). The teacher as applied measurer: Realities of classroom measurement and assessment. *Applied Measurement in Education, 6*(3), 241-254. Retrieved February 7, 2009, from ERIC database.

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House, 78*(5), 218-223.

Allen, J. D. & Lambating, J. (2001). Validity and reliability in assessment and grading: Perspectives of preservice and inservice teachers and teacher education professors. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association (1990). Standards for Teacher Competence in Educational Assessment of Students. *Educational Measurement: Issues and Practice, 9* (4), 30-32.

Bailey, J. & McTighe, J. (1996). Reporting achievement at the secondary level: What and how. In T. R. Guskey (Ed.), *ASCD Yearbook – 1996: Communicating Student Learning* (pp. 119-140). Alexandria, VA: Association for Supervision and Curriculum Development.

Barnes, S. (1985). A study of classroom pupil evaluation: The missing link in teacher education. *The Journal of Teacher Education, 36*(4), 46-49.

Baron, P. A. B. (2000). Consequential validity for high school grades: What is the meaning of grades for senders and receivers. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Black, P. (2000). Research and the development of educational assessment. *Oxford Review of Education, 26*(3&4), 407-419.

Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-76. Retrieved May 4, 2009 from ERIC database.

Black, P. & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Black, P. & Wiliam, D. (2003). "In praise of educational research": Formative Assessment. *British Education Research Journal, 29*(5), 623-637.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123-142.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education, 7*(4), 279-301.

Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J.H. McMillan (Ed.), *Formative Classroom Assessment* (pp. 43-62). Danvers, MA: Teachers College, Columbia University.

Brookhart, S. M. (2008). *How to give effective feedback to your students*. Alexandria, Virginia: Association for Supervision and Curriculum Development (ASCD).

Canady, R. L. & Hotchkiss, P. R. (1989). It's a good score! Just a bad grade. *Phi Delta Kappan, 71*(1), 68-71.

Carlson, L. A. (2003). Beyond assessment to best grading practice: Practical Guidelines. *Measuring up: Assessment issues for teachers, counselors, and administrators.* Retrieved July 21, 2008, from ERIC database.

Carey, L. M. (1988). *Measuring and evaluating school learning.* Boston, Massachusetts: Allyn and Bacon, Inc.

Cauley, K. M., Abrams, L., Pannozzo, G., McMillan, J., & Hearn, J. (2008). *The relationship between classroom assessment practices and student motivation and engagement.* Richmond, VA: Virginia Commonwealth University, Metropolitan Educational Research Consortium.

Cauley, K. M., Pannozzo, G., Abrams, L., McMillan, J., & Camou-Linkroum, S. (2006). *The relationship between classroom assessment practices and student motivation and engagement: A literature review.* Richmond, VA: Virginia Commonwealth University, Metropolitan Educational Research Consortium.

Cross, L. H. & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education, 12*(1), 53-72.

Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement*. Englewood Cliffs, New Jersey: Prentice Hall.

Frary, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice, 12*(3), 23-30.

Friedman, S. J. (1998). Grading teachers' grading policies. *NASSP Bulletin, 82*(597), 77-83. Retrieved July 21, 2008, from http://bul.sagepub.com/cgi/content/abstract/82/597/77

Friedman, S. J. & Frisbie, D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational and Psychological Measurement, 55*(1), pp. 5-26.

Frisbie, D. A. & Waltman, K. K. (1992, Fall). Developing a personal grading plan. *Educational Measurement: Issues and Practice, 11*(3), 35-42.

Guskey, T. R. (1994). Making the grade: What benefits students? *Educational Leadership, 52*(2), 14-20.

Guskey, T. R. (2000). Grading policies that work against standards and how to fix them. *NASSP Bulletin*. Retrieved July 21, 2008, from http://bul.sagepub.com/cgi/content/abstract/84/620/20

Guskey, T. R. (2001a). Helping standards make the grade. *Educational Leadership, 59*(1), 20-27.

Guskey, T. R. (2001b). High percentages are *not* the same as high standards. *Phi Delta Kappan, 82*(7), 534-536.

Guskey, T. R. (2002a). Computerized gradebooks and the myth of objectivity. *Phi Delta Kappan, 83*(10), 775-781.

Guskey, T. R. (2002b). *Perspectives on grading and reporting: Differences among teachers, students, and parents.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Guskey, T. R. (2004). *Stability and change in high school grades.* Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Guskey, T. R. (2007, Spring). Multiple sources of evidence: An analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues and Practice,* 19-27. Retrieved September 28, 2008 from ERIC database.

Hirschenbaum, H., Simon, S. B., & Napier, R. W. (1971). *Wad-ja-get? The grading game in American education.* New York: Hart.

Howley, A., Kusimo, P. S., & Parrott, L. (2001). Grading and the ethos of effort. *Learning Environments Research*, 3(3), 229-246. Retrieved November 23, 2007 from ERIC database.

Huck, S. W. (2000). *Reading statistics and research* (3rd ed.). New York: Addison Wesley Longman, Inc.

Lambating, J. & Allen, J. D. (2002). *How the multiple functions of grades influence their validity and value as measures of academic achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

McMillan, J. H. (1999). The devastating effect of zeros on grades: What can be done? Opinion paper. Richmond, VA: Virginia Commonwealth University.

McMillan, J. H. (2001, Spring). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*. Retrieved September 28, 2008 from ERIC database.

McMillan, J. H. & Nash, S. (2000). Teacher classroom assessment and grading practices decision making. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

McMillan, J. H. & Workman, D. J. (1998). *Classroom assessment and grading practices: A review of the literature*. Richmond, VA: Virginia Commonwealth University, Metropolitan Educational Research Consortium. Retrieved July 21, 2008 from ERIC database.

McMillan, J. H. & Workman, D. J. (1999). *Teachers' classroom assessment and grading practices: Phase I and II.* Richmond, VA: Virginia Commonwealth University, Metropolitan Educational Research Consortium. Retrieved July 21, 2008 from ERIC database.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: Macmillan.

O'Connor, K. (1995). Guidelines for grading that support learning and student success. *NASSP Bulletin, 79*(571), 91-101. Retrieved July 21, 2008, from http://bul.sagepub.com/cgi/content/abstract/79/571/91

Ornstein, A. C. (1994). Grading practices and policies: An overview and some suggestions. *NASSP Bulletin, 78*(561), 55-64. Retrieved July 21, 2008, from http://bul.sagepub.com/cgi/content/abstract/78/561/55

Pilcher, J. K. (1994). The value-driven meaning of grades. *Educational Assessment, 2*(1), 69-88.

Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice, 10*(1), 7-12.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment *for* learning. *Phi Delta Kappan, 83*(10), 758-765.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice, 8*(2), 5-14.

Stiggins, R. & Knight, T. (1997). *But are they learning? A commonsense parents' guide to assessment and grading in schools.* Portland, OR: Assessment Training Institute.

Terwilliger, J. S. (1966). Self-reported marking practices and policies in public secondary schools. *NASSP Bulletin, 50*(5), 5-37. Retrieved July 21, 2008, from http://bul.sagepub.com/cgi/reprint/50/308/5

Terwilliger, J. S. (1989). Classroom standard setting and grading practices. *Educational Measurement: Issues and Practice, 8*(2), 15-19.

Wendel, F. C. & Anderson, K. E. (1994). Grading and marking systems: What are the practices, standards? *NASSP Bulletin, 78*(558), 79-84. Retrieved July 21, 2008, from http://bul.sagepub.com/cgi/content/abstract/78/558/79

Wiliam, D. & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537-548.

Wiliam, D., Lee, C., Harrison, C. & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*(1), 49-65).

Appendices

APPENDIX A

Standards for Teacher Competence in Educational Assessment of Students	
1.	Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
2.	Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
3.	Teachers should be skilled in administering, scoring, and interpreting the results of both externally-produced and teacher-produced assessment methods.
4.	Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
5.	Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.
6.	Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
7.	Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Source: American Federation of Teachers, National Council on Measurement in Education, & National Education Association (1990).

APPENDIX B

Assessment and Grading Practices of Middle School, Core Subject Teachers

Title: Interpreting the Meaning of Grades: An Analysis of Middle School Teachers' Assessment and Grading Practices

VCU IRB NO.: HM12503

Investigators: Dr. Lisa M. Abrams and Tameshia Vaden Grimes, Doctoral Candidate

You have been invited to participate in the research study Interpreting the Meaning of Grades: An Analysis of Middle School Teachers' Assessment and Grading Practices.

The purpose of this study is to understand teachers' grading and assessment practices and how these practices may relate to improved student achievement. You have been asked to participate because you taught a middle school, core academic subject during the 2008-2009 school year.

Description of the Study and Your Involvement
If you decide to participate in this research study, you will be asked to complete the web-based survey that will take approximately 20 minutes to complete. All information will be kept strictly confidential. Survey participants will be assigned a unique numerical code which will be used to link your survey responses with the 2008-2009 average student SOL scores. The data linking will be conducted by the researchers; your participation in the survey will be completely confidential and will not be shared with district personnel. Once the survey and student achievement data has been linked, all potentially identifying information will be removed.

Risks and Discomforts
There are no minimal risks or discomforts associated with this study. You may choose to skip survey questions you prefer not to answer or stop participating in the study at any time.

Benefits to You and Others
You may not derive any direct benefit from this study; however, the information obtained from the study may provide the school district and the research community with important information about how assessment and grading practices impact student achievement.

Costs
There are no costs for participating in this study other than the time you will spend completing the online survey. The survey should take approximately 20 minutes to complete.

Confidentiality
Participation in this study and all survey responses and student achievement data is confidential. Potentially identifiable information includes the unique identifier that will be used to link your survey responses and average student SOL data. Once the linking has been completed this numeric identifier will be removed from study data files. Aggregate data from this study may be used in publications and presentations, but it will not be possible to identify individual participants.

Voluntary Participation and Withdrawal
You do not have to participate in this study. If you choose to paraticipate, you may stop at any time without penalty. You may also choose not to answer certain questions within the survey.

Questions
If you have questions or concerns about the research study please contact Dr. Lisa Abrams or Tameshia Grimes at lmabrams@vcu.edu or grimestv@vcu.edu.

If you have questions regarding your rights as a participant in this study, you may contact the Office for Research, VCU, 800 East Leigh Street, Suite 113, PO Box 980568, Richmond, Virginia 23298 (Telephone 804-827-2157).

Consent
By selecting the "next" button below you are acknowledging that you have read and understand the information about the study and you are providing your consent to participate in the research study.

Please click "Next" to start the survey or select "Decline" to indicate that you prefer not to participate in the study.

This is a confidential survey. Your responses will not be shared with anyone.

The purpose of this questionnaire is to gather information about the assessment and grading practices of middle school, core-subject teachers. For this survey, the term "assessment" refers to any activity designed with the purpose of determining how well students are progressing towards goals and/or objectives; "grading" refers to any marks or scores used to represent how a student is performing.

As you respond to each statement below, please do so with regard to only one section of the course you taught during the 2008-2009 school year, referencing the same section for all responses. There are no right or wrong answers, and all responses will be kept confidential.

Part I – Assessing and Grading

Assessing student progress and grading students' works are major functions of the job of teaching.

Indicate your level of agreement to the question below:

1. Grading/evaluating students is the most difficult aspect of teaching.
{Choose one}
( ) Strongly Disagree
( ) Disagree
( ) Neither Agree nor Disagree
( ) Agree
( ) Strongly Agree

2. I use a variety of assessments to determine how much progress students are making in my class.
{Choose one}
( ) Yes
( ) No

Part II - Course Information

Directions: Identify and select on section of one course you taught during the 2008-2009 school year. Answer the questions below – and the questions in each subsequent section - based on that specific section and course.

3. What is the subject area of the class you referenced when answering the survey questions?
{Choose one}

( ) English
( ) Mathematics
( ) Science
( ) Social Studies

4. What is the primary grade level of the class you referenced when answering the survey questions?
{Choose one}

( ) 6
( ) 7
( ) 8

5. What is the primary ability level for the class you referenced when answering the survey questions?
{Choose one}

( ) Comprehensive
( ) Honors and/or Gifted
( ) Inclusive/Collaborative

6. During what period did the class you referenced when answering the survey questions meet?
{Choose one}

( ) 1st period or 1st period odd
( ) 2nd period or 2nd period odd
( ) 3rd period or 3rd period odd
( ) 4th period or 4th period odd
( ) 5th period or 1st period even
( ) 6th period or 2nd period even
( ) 7th period or 3rd period even
( ) 8th period or 4th period even

Part III - Grading Practices

For question #7, click on the square beside your choice and a number will appear. Continue to click until all squares have a number. If you change your mind about the order of importance, simply re-click in the square to remove the number and then select your choice. The first item you select should be the one that holds the most importance to you.

7. Please select and rank from the choices below what you believe to be the <u>five</u> main purposes of grading (with 1 being the primary purpose for grading).
{Rank the following from 1 to 5}

[ ] Communicate to parents
[ ] Communicate to school personnel
[ ] Evaluate school programs
[ ] Provide feedback to students
[ ] Measure student's level of effort
[ ] Measure student's level of responsibility
[ ] Measure student's mastery of academic content
[ ] Provide incentive to student/motivation
[ ] Select, identify, or group students (within the classroom)

Part III Continued - Grading Practices

Thinking of the same section and course you referenced in questions 3 – 6 above, please indicate your level of agreement with the following statements on grades and/or grading.

8. Students should receive zeros for incomplete assignments.
{Choose one}
( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

9. All assignments should be graded.
{Choose one}
( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

10. All graded assignments should count towards the course grade.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

11. Grades should be used to teach students lessons about responsibility.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

12. Grades should be used to reward students.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

13. Grades should be used to motivate students.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

14. Homework completion should be included when determining students' grades.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

15. Homework accuracy should be included when determining students' grades.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

16. Grades should reflect the student's level of mastery of the content.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

17. Grades should reflect the student's progress/improvement over time.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

18. Grades should reflect the student's ability to follow directions.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

19. Grades should reflect the student's effort.
{ Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

20. Criterion-referenced grading is a legitimate method of identifying a student's ability in my class.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

21. Norm-referenced grading is a legitimate method of identifying a student's ability in my class.
{Choose one}

( ) Strongly Disagree
( ) Disagree
( ) Not Sure
( ) Agree
( ) Strongly Agree

Thinking of the same section and course you used to answer questions 8 - 21 above, please select the option that best describes how much influence the following factors have when you are determining the end-of-course grades students received for your class.

22. Attendance in class.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

23. Behavior/attitude in class.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

24. Extra credit for academic performance.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

25. Extra credit for non-academic performance.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

26. Improvement of performance since the beginning of the year.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

27. Student effort - how much the student tried to learn.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

28. Participation in class.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

29. Performance compared to other students in the class.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

30. Work habits (e.g. neatness, proper format, etc.)
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

31. Classwork
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

32. Daily homework.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

33. Inclusion of zeros for incomplete or missing assignments.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

34. Projects (e.g. posters, timelines, creations, etc.).
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

35. Quiz scores.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

36. Specific learning objectives mastered.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

37. Test scores.
{Choose one}

( ) No influence at all
( ) Minimum influence
( ) Some influence
( ) Quite a bit of influence
( ) Extensive influence

PART IV - Assessment Practices
Directions: For question #38, click on the square beside your choice and a number will appear. Continue to click until all squares have a number. If you change your mind about the order of importance, simply re-click in the square to remove the number and then select your choice. The first item you select should be the one that holds the most importance to you.

38. Please select and rank from the choices below what you believe to be the <u>top five</u> assessment practices that could be used to determine what students know and can do (with 1 being the best indicator of student achievement).
{Rank the following from 1 to 5}

[ ] District assessments/benchmark tests
[ ] End-of-lesson review (independent)
[ ] End-of-lesson review (teacher-led)
[ ] Group work
[ ] Homework completion and quality
[ ] Pre-tests/Post-tests (unit)
[ ] Portfolios of students' work
[ ] State assessments (SOL tests)
[ ] Students' behavior and attitude in class
[ ] Students' class involvement/participation
[ ] Students' exhibits (projects and reports)
[ ] Students' grades on assignments
[ ] Students' self-assessments
[ ] Teacher-developed assessments
[ ] Teacher questioning and observations

Part IV Continued - Assessment Practices

Thinking of the same section and course you used to answer questions 22 - 37 above, please select the frequency with which you used the following assessment practices to determine students' achievement levels (what students know and can do).

39. District assessments
{Choose one}

( ) Never
( ) Seldom (1-2 times/year)
( ) Occasionally (3-4 times/year)
( ) Frequently (5-6 times/year)
( ) Always

40. State assessments (SOLs)
{Choose one}

( ) Never
( ) Seldom (1-2 times/year)
( ) Occasionally (3-4 times/year)
( ) Frequently (5-6 times/year)
( ) Always

41. End-of-lesson review (independent)
{Choose one}

( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

42. Group work
{Choose one}

( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

43. Homework completion and quality
{Choose one}

( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

44. Pre-tests/post-tests
{Choose one}

( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

45. Student exhibits (projects and reports)
{Choose one}
( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

46. Students' self-assessment
{Choose one}
( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

47. Students' behavior/attitude in class
{Choose one}
( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

48. Students' class involvement/participation
{Choose one}
( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

49. Students' grades
{Choose one}
( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

50. Teacher questioning and observation (in class)
{Choose one}

( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

51. Teacher-developed assessments
{Choose one}

( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

Thinking of the same section and course you used to answer questions 39 - 51 above, please select the frequency with which you used the following types of questions or assessment methods to determine students' mastery of the learning objectives (what students know and can do).

52. Essays
{Choose one}

( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

53. Matching
{Choose one}

( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

54. Multiple choice
{Choose one}

( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

55. Short answer
{Choose one}
( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

56. True-False
{Choose one}
( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

57. Student exhibits (projects and reports)
{Choose one}
( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

58. Portfolios of students' work
{Choose one}
( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

59. Other (please specify below):
{Choose one}
( ) Never
( ) Seldom (quarterly)
( ) Occasionally (monthly)
( ) Frequently (weekly)
( ) Always (daily)

59b. If you selected #59 above, please list the other types of test questions you used during the 2008-2009 school year with the section and course you thought of while completing this survey.
{Enter answer in paragraph form}
[

]

Part V - Grading Policies and Grading Plans

Directions: Please type your answers to the following questions in the box provided immediately beneath each question.

60. Does your school have a specific policy related to grades/grading that you are required to follow?
{Choose one}
( ) Yes
( ) No

60b. If you answered yes to question #60 above, please describe your school's grading policy in the space below:
{Enter answer in paragraph form}
[

]

61. Have you developed a personal grading plan/policy that guides your approach to grading?
{Choose one}
( ) Yes
( ) No

61b. If you answered yes to question #61, have you shared your plan with other teachers?
{Choose one}
( ) Yes
( ) No

61c. Describe your personal grading plan/policy in the space provided below:
{Enter answer in paragraph form}
[

]

PART VI - Participant Information

Directions: Answer the following questions as they relate specifically to you and your experience as a classroom teacher.

62. What is your gender?
{Choose one}
( ) Female
( ) Male

63. What is your ethnicity/racial background?
{Choose one}
( ) American Indian
( ) Asian
( ) Black/African American
( ) Hispanic
( ) White/Caucasian
( ) Hawaiian/Pacific Islander
( ) Other

64. What is your age group?
{Choose one}
( ) 20-25
( ) 26-30
( ) 31-35
( ) 36-40
( ) 41-45
( ) 46-50
( ) 51-55
( ) 56-60
( ) 61+

65. Which choice below best describes your endorsement area?
{Choose one}

( ) Elementary Endorsement (grades PreK – 6)
( ) Middle School Endorsement (grades 6 – 8)
( ) Secondary Endorsement (grades 7 – 12)

66. How many years have you been teaching prior to the 2009-2010 school year?
{Choose one}

( ) 1 - 3 years
( ) 4 - 9 years
( ) 10-14 years
( ) 15+ years

Thank you for your participation!

APPENDIX C

Data Analysis Chart

Research Questions	Instrument Items	Data Analysis
1. What do middle school, core academic subject teachers believe to be the purpose of grades?	7	Descriptive statistics: frequencies overall as well as by subject area, grade level, and student ability level
2. What assessment and grading practices do middle school, core academic subject teachers use most often in their classrooms when determining students' grades?	8-21, 22-37, 38, 39-59	Descriptive statistics: frequencies overall as well as by subject area, grade level, and student ability level
3. Is there a relationship between assessment and grading practices used by middle school, core academic subject teachers and students' SOL scores for the subject area?	22-37, 52-58 End-of-course Average SOL Average	Bivariate Correlation: overall as well as by subject area, grade level, and student ability level

VITA

Tameshia Vaden Grimes was born in 1974 and raised in Franklin, Virginia. She is the oldest of three children born to Ronald M. and Miriam E. Vaden. After graduating, with honors, from Franklin High School, Tameshia attended, and graduated from, Randolph-Macon College in Ashland, Virginia, with a Bachelor's of Arts degree in English and a minor in secondary education. While working full time in the field of human resources, Tameshia earned a Master of Arts degree in English Literature from Virginia Commonwealth University. Although she intended to continue in the field of human resources, Tameshia met the man who would become her husband, and who worked for the same company and in the same division, so she decided to pursue a career in education, became a teacher, and never looked back. Tameshia taught 7th grade Language Arts for five years. While teaching, she earned her third degree, Education Specialist (Ed.S.), in Administration and Supervision from The George Washington University. While completing course work for the Ed.S., Tameshia was part of a group that published the results of a research study: *Assessed Needs of Women at College*, 2001. In 2003, she became an assistant principal, and in 2007 she was appointed principal of a middle school, the position she holds today.